



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification :</b>  <b>Not classified</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 98/38846</b>  <b>(43) International Publication Date:</b> 11 September 1998 (11.09.98)
<b>(21) International Application Number:</b> PCT/US98/04571 <b>(22) International Filing Date:</b> 6 March 1998 (06.03.98)  <b>(30) Priority Data:</b> 08/813,159 7 March 1997 (07.03.97) US 60/042,125 28 March 1997 (28.03.97) US  <b>(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application</b> US 60/042,125 (CIP) Filed on 28 March 1997 (28.03.97)  <b>(71) Applicant (for all designated States except US):</b> AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> LIPSHUTZ, Robert, J. [US/US]; 970 Palo Alto Avenue, Palo Alto, CA 94301 (US). CHEE, Mark [AU/US]; 3199 Waverly Street, Palo Alto, CA 94306 (US). FAN, Jian-Bing [CN/US]; 275 Ventura Avenue #20, Palo Alto, CA 94306 (US). BERNO, Anthony [CA/US]; 570 South 12th Street, San Jose, CA 95112 (US).		<b>(74) Agents:</b> LIEBESCHUETZ, Joe et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US).  <b>(81) Designated States:</b> JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With declaration under Article 17(2)(a). Without classification and without abstract; title not checked by the International Searching Authority.</i>
<b>(54) Title:</b> GENETIC COMPOSITIONS AND METHODS		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## GENETIC COMPOSITIONS AND METHODS

### 5                    CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of USSN 08/813,159, filed March 7, 1997 and USSN 60/042,125 filed March 28, 1997, which are incorporated by reference in their entirety for all purposes.

### BACKGROUND OF THE INVENTION

10                   The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution generating variant forms of progenitor sequences (Gusella, *Ann. Rev. Biochem.* 55, 831-854 (1986)). The variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers a lethal disadvantage and is not transmitted to subsequent generations of the  
15 organism. In other instances, a variant form confers an evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. In many instances, both progenitor and variant form(s) survive and co-exist in a species population. The coexistence of multiple forms of a sequence gives rise to polymorphisms.

20                   Several different types of polymorphism have been reported. A restriction fragment length polymorphism (RFLP) means a variation in DNA sequence that alters the length of a restriction fragment as described in Botstein et al., *Am. J. Hum. Genet.* 32, 314-331 (1980). The restriction fragment length polymorphism may create or delete a restriction site, thus changing the length of the restriction fragment. RFLPs have been widely used in  
25 human and animal genetic analyses (see WO 90/13668; W090/11369; Donis-Keller, *Cell* 51, 319-337 (1987); Lander et al., *Genetics* 121, 85-99 (1989)). When a heritable trait can be linked to a particular RFLP, the presence of the RFLP in an individual can be used to predict the likelihood that the animal will also exhibit the trait.

Other polymorphisms take the form of short tandem repeats (STRs) that include  
30 tandem di-, tri- and tetra-nucleotide repeated motifs. These tandem repeats are also referred to as variable number tandem repeat (VNTR) polymorphisms. VNTRs have been used in identity and paternity analysis (US 5,075,217; Armour et al., *FEBS Lett.* 307, 113-115

(1992); Horn et al., WO 91/14003; Jeffreys, EP 370,719), and in a large number of genetic mapping studies.

Other polymorphisms take the form of single nucleotide variations between individuals of the same species. Such polymorphisms are far more frequent than RFLPs, STRs and VNTRs. Some single nucleotide polymorphisms occur in protein-coding sequences, in which case, one of the polymorphic forms may give rise to the expression of a defective or other variant protein and, potentially, a genetic disease. Examples of genes, in which polymorphisms within coding sequences give rise to genetic disease include  $\beta$ -globin (sickle cell anemia) and CFTR (cystic fibrosis). Other single nucleotide polymorphisms occur in noncoding regions. Some of these polymorphisms may also result in defective protein expression (e.g., as a result of defective splicing). Other single nucleotide polymorphisms have no phenotypic effects.

Single nucleotide polymorphisms can be used in the same manner as RFLPs, and VNTRs but offer several advantages. Single nucleotide polymorphisms occur with greater frequency and are spaced more uniformly throughout the genome than other forms of polymorphism. The greater frequency and uniformity of single nucleotide polymorphisms means that there is a greater probability that such a polymorphism will be found in close proximity to a genetic locus of interest than would be the case for other polymorphisms. Also, the different forms of characterized single nucleotide polymorphisms are often easier to distinguish than other types of polymorphism (e.g., by use of assays employing allele-specific hybridization probes or primers).

Despite the increased amount of nucleotide sequence data being generated in recent years, only a minute proportion of the total repository of polymorphisms in humans and other organisms has so far been identified. The paucity of polymorphisms hitherto identified is due to the large amount of work required for their detection by conventional methods. For example, a conventional approach to identifying polymorphisms might be to sequence the same stretch of oligonucleotides in a population of individuals by didoxy sequencing. In this type of approach, the amount of work increases in proportion to both the length of sequence and the number of individuals in a population and becomes impractical for large stretches of DNA or large numbers of persons.

## SUMMARY OF THE INVENTION

The invention provides nucleic acid segments of between 10 and 100 bases from a fragment shown in Table 1, column 1 including a polymorphic site. Complements of these segments are also included. The segments can be DNA or RNA, and can be double- or  
5 single-stranded. Some segments are 10-20 or 10-50 bases long. Preferred segments include a diallelic polymorphic site. The base occupying the polymorphic site in the segments can be the reference (Table 1, column 3) or an alternative base (Table 1, column 5).

The invention further provides allele-specific oligonucleotides that hybridizes to a segment of a fragment shown in Table 1, column 8 or its complement. These  
10 oligonucleotides can be probes or primers. Also provided are isolated nucleic acids comprising a sequence of Table 1, column 8, or the complement thereto, in which the polymorphic site within the sequence is occupied by a base other than the reference base shown in Table 1, column 3.

The invention further provides a method of analyzing a nucleic acid from an  
15 individual. The method determines which base is present at any one of the polymorphic sites shown in Table 1. Optionally, a set of bases occupying a set of the polymorphic sites shown in Table 1 is determined. This type of analysis can be performed on a plurality of individuals who are tested for the presence of a disease phenotype. The presence or absence of disease phenotype can then be correlated with a base or set of bases present at the polymorphic sites  
20 in the individuals tested.

## DEFINITIONS

An oligonucleotide can be DNA or RNA, and single- or double-stranded. Oligonucleotides can be naturally occurring or synthetic, but are typically prepared by synthetic means. Preferred oligonucleotides of the invention include segments of DNA, or  
25 their complements including any one of the polymorphic sites shown in Table 1. The segments are usually between 5 and 100 bases, and often between 5-10, 5-20, 10-20, 10-50, 20-50 or 20-100 bases. The polymorphic site can occur within any position of the segment. The segments can be from any of the allelic forms of DNA shown in Table 1.

Hybridization probes are oligonucleotides capable of binding in a base-specific  
30 manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids,

as described in Nielsen et al., *Science* 254, 1497-1500 (1991).

The term primer refers to a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (*i.e.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair means a set of primers including a 5' upstream primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3', downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

Linkage describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome, and can be measured by percent recombination between the two genes, alleles, loci or genetic markers.

Polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphic locus may be as small as one base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms.

A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually

preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations).

A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine  
5 by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25 °C. For example,  
10 conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C are suitable for allele-specific probe hybridizations.

An isolated nucleic acid means an object species invention that is the predominant species present (*i.e.*, on a molar basis it is more abundant than any other individual species in the composition). Preferably, an isolated nucleic acid comprises at least about 50, 80 or  
15 90 percent (on a molar basis) of all macromolecular species present. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods).

Linkage disequilibrium or allelic association means the preferential association of a particular allele or genetic marker with a specific allele, or genetic marker at a nearby  
20 chromosomal location more frequently than expected by chance for any particular allele frequency in the population. For example, if locus X has alleles a and b, which occur equally frequently, and linked locus Y has alleles c and d, which occur equally frequently, one would expect the combination ac to occur with a frequency of 0.25. If ac occurs more frequently, then alleles a and c are in linkage disequilibrium. Linkage disequilibrium may result from  
25 natural selection of certain combination of alleles or because an allele has been introduced into a population too recently to have reached equilibrium with linked alleles.

A marker in linkage disequilibrium can be particularly useful in detecting susceptibility to disease (or other phenotype) notwithstanding that the marker does not cause the disease. For example, a marker (X) that is not itself a causative element of a disease, but which is in  
30 linkage disequilibrium with a gene (including regulatory sequences) (Y) that is a causative element of a phenotype, can be used detected to indicate susceptibility to the disease in

circumstances in which the gene Y may not have been identified or may not be readily detectable.

The present invention includes the use of any of the polymorphic forms shown in Table 1 as a means to determine susceptibility to a phenotype resulting from an allele or marker in  
5 linkage disequilibrium with such polymorphic forms.

## DESCRIPTION OF THE PRESENT INVENTION

### I. Novel Polymorphisms of the Invention

The novel polymorphisms of the invention are listed in Table 1. The first column of the Table lists the names assigned to the fragments in which the polymorphisms occur.  
10 The fragments are all human genomic fragments. SGC, TIGR and WI respectively stand for Stanford Genome Center, The Institute for Genome Research and the Whitehead Institute. The sequence of one allelic form of each of the fragments (arbitrarily referred to as the prototypical or reference form) has been previously been determined. Many of these sequences are listed at <http://www-genome.wi.mit.edu/>; <http://shgc.stanford.edu>; or  
15 <http://www.tigr.org/>. The Web sites also list primers for amplification of the fragments, and the genomic location of fragments. Some fragments are expressed sequence tags, and some are random genomic fragments. All information in the websites concerning the fragments listed in Table 1 is incorporated by reference in its entirety for all purposes.

The second column lists the position in the fragment in which a polymorphic site has  
20 been found. Positions are numbered consecutively with the first base of the fragment sequence as listed in one of the above databases being assigned the number one. The third column lists the base occupying the polymorphic site in the sequence in the data base. This base is arbitrarily designated the reference or prototypical form but is not necessarily the most frequently occurring form. The fifth column in the table lists the alternative base(s) at the  
25 polymorphic site. The eighth column of the Table lists about 15 bases of sequence on either side of the polymorphic site in each fragment. The indicated sequences can be either DNA or RNA. In the latter, the T's shown in the Table are replaced by U's. The base occupying the polymorphic site is indicated in EUPAC-IUB ambiguity code. The fourth and sixth columns of the table show the frequency with which reference and alternative alleles occur  
30 at a polymorphic site. The seventh column in the table indicates the population frequency of heterozygotes of the polymorphic site.



Table 1

Fragment	Position	Ref. Allele	Frequency (p)	Alt. Allele	Frequency (q)	Heterozygosity (h)	Sequence tag
SGC35469	118 A		0.75 C		0.25	0.38	TTAAGTGAGAMTCITTTAAAC
SGC35512	34 T		0.5 C		0.5	0.5	AGAGCCGTCYCTCAGGTTGC
SGC35512	50 G		0.31 C		0.69	0.43	GTGCGCTGTCSTCTCCTGGCC
SGC35594	74 C		0.63 G		0.37	0.47	GGCGCATCCSTTAGTTTCCA
SGC35681	42 T		0.5 C		0.5	0.5	AGAGAAAAAAYCAACAGCAAA
SGC35681	56 A		0.56 C		0.44	0.49	CAGCAAAACAAMACCACACAAA
SGC35683	34 T		0.75 G		0.25	0.38	CAATAAGCACKCATGACCTCA
TIGR-A003N21	49 C		0.94 A		0.06	0.12	GTGATTTGGTMAGCATATCTT
TIGR-A004S25	145 G		0.79 A		0.21	0.34	TGTACTTTGGRCTCCAGACTT
TIGR-A004V30	203 C		0.67 G		0.33	0.44	AGTAGAAAAAGSCTTCTAGGTT
TIGR-A004W22	232 C		0.92 A		0.08	0.15	CCCCCGCCTAMCIGGAGATGT
TIGR-A004Z48	177 A		0.38 G		0.62	0.47	ACGCCACAGARTCCTCCAATT
TIGR-A005D24	123 A		0.94 G		0.06	0.12	ATAGAGAAATRAAAACCCAAT
TIGR-A005D24	138 C		0.75 T		0.25	0.38	CCCAATTTCTYTTTCACCAAT
WI-10072	105 G		0.83 A		0.17	0.28	TATTTTGTGTGACTCTAT
WI-10088	205 C		0.86 G		0.14	0.24	TTTAGACAGGSAGCAGAAGCA
WI-1017	93 G		0.57 A		0.43	0.49	ACCAGACAAGRGATGTAGATT
WI-1021	24 A		0.69 T		0.31	0.43	ATCAAAGCACWATCTGTGTTT
WI-1031	149 G		0.75 A		0.25	0.38	GATGCCAGCARGACACACACCC
WI-10396	72 C		0.29 A		0.71	0.41	TGGGAAGAGTMTGTGACTTTA
WI-10400	46 T		0.43 C		0.57	0.49	TAGAAAGTAAYTGCAATTCAG
WI-10400	165 A		0.86 T		0.14	0.24	CTCCCCACCWAAAAATAACGT
WI-10400	166 A		0.86 T		0.14	0.24	TCCCCACCCAWAAAAATAACGTA
WI-10400	189 A		0.43 G		0.57	0.49	TACCTATGTCRTGCCATGTAG
WI-10613	44 G		0.19 A		0.81	0.3	GAAACATACARTGTAATAGAA
WI-10613	172 A		0.06 C		0.94	0.12	ATTTTATTGMCCTCCTAGGAG
WI-10616	116 G		0.94 C		0.06	0.12	GTAGGTCCTGCTCCTATCA

WI-10616	141 C	0.5 T	0.5	0.5	0.5	GCCACGTAGCYCTCCCTCCC
WI-10656	59 T	0.5 G	0.5	0.5	0.5	TTCTTTTGGCTCTTAGAAT
WI-10673	94 C	0.69 G	0.31	0.31	0.43	ATGGAGGGGGSTGCAGGTGG
WI-10681	41 A	0.58 T	0.42	0.42	0.49	GACCCCATTTGWCCTTACGCAA
WI-10681	103 T	0.58 A	0.42	0.42	0.49	TAAAAAAGCCWAAAGACAGCC
WI-10685	25 A	0.86 G	0.14	0.14	0.24	TGGATAGGTCRACCGGCTGAA
WI-10744	61 G	0.33 C	0.67	0.67	0.44	ACAAAAGGACSAAAAACACTC
WI-10770	49 G	0.71 T	0.29	0.29	0.41	CTCCCTTCTKCCTGGCCCTT
WI-10770	174 G	0.64 A	0.36	0.36	0.46	AGGACACTCARITTCACATGCC
WI-10809	33 C	0.71 T	0.29	0.29	0.41	AAACCATGAAYGGTATAAGGA
WI-10809	78 C	0.57 T	0.43	0.43	0.49	CCTCTCACCAYTTAGAAAAGG
WI-10826	132 A	0.71 C	0.29	0.29	0.41	AAGACCTGCAMCCCTGGCTTC
WI-10854	102 C	0.33 T	0.67	0.67	0.44	AGTTGAAACAYGAAGACGATA
WI-10854	152 G	0.33 T	0.67	0.67	0.44	CGAGGCAACAKGGAGAGGTAC
WI-10870	103 G	0.69 A	0.31	0.31	0.43	CCTACTTAGARCAGTGGAGTA
WI-11152	179 C	0.06 T	0.94	0.94	0.12	AGGCTTGTCAYCTGTCAGAAA
WI-11183	118 C	0.58 T	0.42	0.42	0.49	GTATTTTGCYCTTGTCACATA
WI-11183	124 C	0.83 T	0.17	0.17	0.28	TTGCCCTTGTACTAACATTT
WI-11183	192 T	0.83 C	0.17	0.17	0.28	GAGTTTAAAYATTGGTATGT
WI-1126	97 T	0.13 C	0.87	0.87	0.22	TTTTCAAGATYCAATATATAT
WI-1126	230 T	0.63 C	0.37	0.37	0.47	GTAACCTTTTGGAGTTGTCT
WI-1795	47 T	0.38 C	0.62	0.62	0.47	ATGCTGGGTYCTTCCAGACT
WI-1795	130 T	0.38 C	0.62	0.62	0.47	GAAAGAAAAGYCGTCTACCAT
WI-1819	51 C	0.94 T	0.06	0.06	0.12	CTTTCAGCACYTTGTTGGATC
WI-1936	117 T	0.58 C	0.42	0.42	0.49	TTGTATCACCYCTCCCGCAAC
WI-1968	167 A	0.75 G	0.25	0.25	0.38	TGGAAGTTGTRTGAACCTGAG
WI-2529	71 C	0.06 T	0.94	0.94	0.12	AGCCTCTCAAYTCTTAACTGC
WI-3429	62 C	0.5 T	0.5	0.5	0.5	GGGCTCCACAYAGCCCTCAGC
WI-3429	64 G	0.44 T	0.56	0.56	0.49	GCTCCACACAKCCCTCAGCCC
WI-3678	125 G	0.69 T	0.31	0.31	0.43	TGATGCACCTKCCTTTTGAT
WI-4582	226 T	0.94 C	0.06	0.06	0.12	AAAATATGGTYCCTCCTTGCT
WI-4687	121 G	0.43 T	0.57	0.57	0.49	AAGGGCACTTKGAGGAGTGT

WI-4701	198 G	0.25 A	0.75	0.38	CCCAATTAGARCCATGTCAAT
WI-4719	70 G	0.56 A	0.44	0.49	AGCGGATTATRTCTGACGCCA
WI-4767	50 A	0.33 G	0.67	0.44	CTTAGACTGARATTCATAAAG
WI-4767	173 C	0.83 A	0.17	0.28	AGGGATGACAMAAATCACTAA
WI-4823	164 C	0.5 A	0.5	0.5	ATTCTAAAMAAAGAAAAGT
WI-4860	72 A	0.71 G	0.29	0.41	TGCTTGATTTGGAGATAAAA
WI-5222	52 G	0.29 C	0.71	0.41	CTCCATCCTASGATTCGCT
WI-5381	178 A	0.63 T	0.37	0.47	TTAGTTTGTWTTACTAAAC
WI-5385	110 G	0.67 A	0.33	0.44	CCAGGAATCGRCAATGCTAAT
WI-563	87 G	0.75 A	0.25	0.38	GGCCTCCCTRCCCTGATCAT
WI-5696	61 C	0.07 A	0.93	0.13	CCTTAGTTTCMTAAAGCCCC
WI-5760	187 G	0.5 A	0.5	0.5	TTAGATAAGCRTCCACGAAA
WI-5801	48 A	0.25 G	0.75	0.38	GTGCTTTGTRGAATTTGAAA
WI-5801	157 G	0.25 A	0.75	0.38	AGCTGGGAARAGGGAATGAG
WI-5826	134 T	0.67 C	0.33	0.44	TATTCCTTAGYTTTCAAAATTA
WI-5865	99 T	0.43 A	0.57	0.49	TATCAAAAATWAAACAAATAT
WI-5865	103 C	0.86 G	0.14	0.24	AAAAATTAASAAATATTAAT
WI-5865	165 T	0.57 A	0.43	0.49	CAAGACACAGWCCAGTCTCCA
WI-5967	148 C	0.92 T	0.08	0.15	ATGCTTGGTAYTTGCTCTGTG
WI-5967	165 C	0.75 T	0.25	0.38	TGTGCCGTATYTGCTCCAATC
WI-6093	53 G	0.88 C	0.12	0.22	CTTTGGCCASGTCTGTAATG
WI-6190	165 G	0.5 A	0.5	0.5	GAGGATCTTGRGAAGCAGCAG
WI-6213	164 C	0.94 G	0.06	0.12	TATACTATGTSATATAATAAT
WI-6238	175 G	0.56 A	0.44	0.49	TCTCAAAAATTRGTTCCAGACT
WI-6275	148 G	0.43 C	0.57	0.49	GCTTGGGAAAASGGAAGGAAAC
WI-6315	187 T	0.75 C	0.25	0.38	TTGCTGATAGYAGTGTCTGG
WI-6315	193 C	0.94 T	0.06	0.12	ATAGTAGTGTCTGGTTCTTC
WI-6554	195 C	0.86 G	0.14	0.24	GAGAGAAAACSCCTGACTTCA
WI-6644	134 T	0.92 A	0.08	0.15	CTCAAGCACAWACCCAAACTT
WI-6711	36 T	0.75 C	0.25	0.38	GACTCCAAAAAYTGAATAAGTA

WI-6711	226 G	0.88 T	0.12	0.22	CACACCCACAKTGGCAACTAA
WI-6786	106 A	0.67 T	0.33	0.44	CTTTGGCGAAGGATAAGAA
WI-6786	111 A	0.5 T	0.5	0.5	GCAGAAAGGATWAAGAAGTGAG
WI-6786	151 G	0.58 A	0.42	0.49	CCATTCTTCTRTGGGATAAGG
WI-6824	112 A	0.88 G	0.12	0.22	GTGTGCCAARACACCTTAGAA
WI-6844	225 T	0.75 C	0.25	0.38	GTCTTGAGGTYATCATTATGA
WI-6905	215 T	0.75 A	0.25	0.38	ACATGAAAAAWAGAGCCTAAG
WI-6911	216 T	0.88 C	0.12	0.22	TTTACCACCTTYCATGACATTG
WI-6962	78 A	0.63 G	0.37	0.47	GATCCAGAGARGACAAAGCTC
WI-7008	180 A	0.31 G	0.69	0.43	CTCTCAAAAGRAGAGTAGTTA
WI-7023	56 A	0.38 C	0.62	0.47	TTTGTGACAGMCCCTGCGTGC
WI-7023	206 C	0.31 A	0.69	0.43	ATTCAACACAMACACACATTG
WI-7038	31 G	0.69 A	0.31	0.43	GGACCTTGGCRCTCTCAGCTT
WI-7038	140 A	0.63 C	0.37	0.47	CCAGACAAAGAMGACTGTCAGG
WI-7038	266 T	0.56 C	0.44	0.49	GAGACTTTTCYCGGTGATGGC
WI-7041	174 C	0.56 A	0.44	0.49	CTGCTCTCMCCACCTTCTT
WI-7069	93 G	0.13 A	0.87	0.22	TTAACAGAGTRTCAGATCTAT
WI-7070	226 C	0.94 T	0.06	0.12	ATGGTGCTTYYAGTTTAATGC
WI-7079	293 T	0.31 G	0.69	0.43	AGATGAAATTKATTTCCATCT
WI-7093	54 C	0.88 T	0.12	0.22	GCCCTTCCCTYGGCTCCCAGC
WI-7104	167 C	0.5 A	0.5	0.5	AGCATGAGGCMCAGCAAGAAG
WI-7104	249 C	0.56 T	0.44	0.49	AGCATCTTTGYTGGCAGGGC
WI-7166	59 C	0.94 T	0.06	0.12	ATCAGTTCTAYGGATCATCAA
WI-7222	126 G	0.69 T	0.31	0.43	GGGGATGGGKAATAAAGGAG
WI-7222	255 G	0.69 A	0.31	0.43	CATTTTCTCARTCATTTCTCT
WI-7224	134 T	0.94 C	0.06	0.12	TGTCAGCATTYATTAAAAAAC
WI-7227	24 A	0.88 G	0.12	0.22	CTCCTGGAGGRAGCCCAGGCA
WI-7227	93 G	0.5 T	0.5	0.5	TTTCAGACAAKCTTTAGAGAA
WI-7227	99 G	0.5 C	0.5	0.5	ACAAGCTTTASAGAAATGGAC
WI-7227	291 G	0.69 A	0.31	0.43	TAAGGGTTGARCAGTTAAAC

WI-7259	189 T	0.44 C	0.56	0.49	CTGGCCACAGYTGCGGAGCA
WI-7307	128 G	0.69 T	0.31	0.43	CCTCCCTCAGKAACTGGAGGA
WI-7310	64 T	0.13 A	0.87	0.22	ACAAGGAACCCWCCGAAGAGGA
WI-7310	234 A	0.44 C	0.56	0.49	CCCCATCCCAMATGATCTTGA
WI-7313	256 C	0.25 T	0.75	0.38	TAGCGATGACYTCTTAATTAT
WI-7313	266 T	0.25 C	0.75	0.38	CTCTTAATTAYAAATTTGATTT
WI-7322	275 A	0.5 G	0.5	0.5	ATAACAGAAATRACTTGCCATC
WI-7330	207 C	0.5 T	0.5	0.5	AAAGTGAGAGYTGAAAAGAGA
WI-7381	54 C	0.25 G	0.75	0.38	GGGAAATCCSCITTTCTTCT
WI-7381	213 C	0.56 T	0.44	0.49	AAACGGCCTCYGGCTCTCAGA
WI-7416	137 G	0.06 T	0.94	0.12	TGGCAGTGCTKCTACTCCTCA
WI-7461	153 C	0.88 T	0.12	0.22	GACTGTGTCTYGTTCCTGTT
WI-7587	28 C	0.56 T	0.44	0.49	AGGTAGCTCCYGAAGATCTGT
WI-7587	81 G	0.5 A	0.5	0.5	TCCCTTCTGRATCTGAAAAG
WI-7587	133 A	0.19 T	0.81	0.3	CCTGAGGAAAWGGAATGAACC
WI-7676	139 C	0.56 T	0.44	0.49	GTGAAGGGGCGGCTTCTCTT
WI-7676	309 A	0.5 C	0.5	0.5	GTGTCCTTGGMAAACTACCTA
WI-7685	46 T	0.13 C	0.87	0.22	TTTTGGGCTCYTTTTTCTCCC
WI-7718	42 A	0.44 C	0.56	0.49	TTACTCAAGCMGTACTCCCT
WI-7718	222 C	0.31 T	0.69	0.43	TTACAAAGAAYCATGCAGGAA
WI-7718	248 A	0.5 G	0.5	0.5	ACTATGTATTTRATTTAGAATG
WI-7719	163 A	0.63 G	0.37	0.47	ACAGTTATCCRTTAGATCAAG
WI-7719	281 T	0.19 C	0.81	0.3	ATCTAGAATCYCTTTATGTTT
WI-7721	145 A	0.75 C	0.25	0.38	CTGTCTCTGCMCTGACTCTC
WI-7805	101 A	0.25 G	0.75	0.38	GAATATGTGTRTGTTAAAGGA
WI-7842	57 T	0.58 C	0.42	0.49	TCCCATTCCTGYGTATGAGTCC
WI-7850	57 G	0.69 A	0.31	0.43	CTGCCTCTGGRCCTCATGTATC
WI-7860	50 C	0.75 T	0.25	0.38	CCTCTCCCCAYTGGGAGAGA
WI-7878	51 C	0.25 G	0.75	0.38	TGATGGCCTGSTGGTTGATAA
WI-7878	162 A	0.19 G	0.81	0.3	GGAGGAGCTGRGTGTGATGAA

WI-7928	101 T	0.14 G	0.86	0.24	TCAAAATTCAKACAAGAGGAA
WI-7933	96 G	0.75 A	0.25	0.38	TTGGCCAGGRCCTCGTATCC
WI-7936	131 T	0.56 A	0.44	0.49	TACACCAAAACWACTGAATGAA
WI-7944	99 T	0.19 C	0.81	0.3	GACTTTCATGYAGCCCCAAAGT
WI-8007	242 C	0.92 A	0.08	0.15	ACTGTTGGACMAGCTGCTGGA
WI-8010	247 G	0.75 T	0.25	0.38	AGTGGTGGGKCTTCCACGTG
WI-8039	87 T	0.94 C	0.06	0.12	TTGTTTCAGTYAAATATGTAT
WI-8039	97 T	0.06 C	0.94	0.12	TAAATATGTAYGTGCCGTGC
WI-8044	107 C	0.58 A	0.42	0.49	GGTTTCTCCCMAGTATGGATT
WI-8053	242 T	0.08 A	0.92	0.15	ACTTATATAAWTTCAGAACTA
WI-8054	131 C	0.63 G	0.37	0.47	CAAGCCTTAGSACAATCTTCT
WI-8054	148 T	0.56 C	0.44	0.49	TTCCTTGTAGYTTTAGCCTTT
WI-8054	237 G	0.5 T	0.5	0.5	GGCGTACAGAKAATCCTTGCC
WI-8057	87 T	0.57 A	0.43	0.49	AAAAGGACAGWGATGGACAGC
WI-8170	204 T	0.88 A	0.12	0.22	CAATCAGAAAWAAAGGTAAAA
WI-8170	259 G	0.56 A	0.44	0.49	ACAAGAAAGCARGCACTTAAAT
WI-8456	93 G	0.38 C	0.62	0.47	GGATGTCACASTTATGTCAAG
WI-8496	41 G	0.79 A	0.21	0.34	GAATGGTAATRTTGTATCAGT
WI-8496	167 A	0.79 G	0.21	0.34	TGCCAATGCARTTAGTATATA
WI-867	119 G	0.56 A	0.44	0.49	TTTCATCTCCRTTGTGIGTT
WI-931	31 A	0.5 G	0.5	0.5	CGGAAGCCACRGCCACTAGCC
WI-931	191 C	0.5 A	0.5	0.5	CAAAAAAGCCMCGAGCCTGGT
WI-9443	211 G	0.81 A	0.19	0.3	CTGACGAGACRCAGAGACCTT
WI-9448	184 G	0.31 A	0.69	0.43	CTGGCACCACRCACCTGGTTTC
WI-9484	178 G	0.92 A	0.08	0.15	GCCAGACAGGRAGGAATTCAA
WI-9617	37 G	0.88 T	0.12	0.22	ACACGCCGTGKTGGCACAGTC
WI-9651	105 A	0.56 T	0.44	0.49	TCGTCTTCAWGGGGCAGCTT
WI-9651	139 T	0.88 C	0.12	0.22	TAGACACCTCYACAGGTACAG
WI-9657	121 T	0.67 G	0.33	0.44	CAAAATAAGKATAATCTTT
WI-9667	68 G	0.81 C	0.19	0.3	TTGTATCATGSTITTATCACTGG

WI-9667	82 C		0.75 T		0.25	0.38	TCACTGGACAYAGCCACCTCC
WI-9702	179 C		0.56 T		0.44	0.49	CAGTTTATTYTAACITTTAAT
WI-9702	344 C		0.5 T		0.5	0.5	AAGACTGGAGYGCCTCAGCCTG
WI-9702	345 G		0.38 A		0.62	0.47	AGACTGGAGCRCTCAGCCTGC
WI-9705	111 C		0.5 A		0.5	0.5	TTGGCTGCCMAAAATTGTTA
WI-9711	390 C		0.5 A		0.5	0.5	GGCATAAGTGMAGGAAAGAGA
WI-9711	423 T		0.69 A		0.31	0.43	AGGAAAAAAWGTATCTGCT
WI-9716	221 G		0.81 A		0.19	0.3	AATTCTAGAAARAAAACACCTA
WI-9760	49 C		0.86 T		0.14	0.24	CTCTCTTTACYAAGTGTTACT
WI-9814	104 C		0.92 T		0.08	0.15	GCTGCTATCTYTTCTCCTTCA
WI-9823	97 C		0.57 T		0.43	0.49	GTGAAATTCYGGGGCATGGG
WI-9825	123 A		0.94 T		0.06	0.12	TCAGGGTCTWGAGGATTAGT
WI-9826	125 A		0.5 T		0.5	0.5	AGAGGCTGTTWTGGCCTTCAA
WI-9826	127 G		0.5 A		0.5	0.5	AGGCTGTTATRGCTTCAAAG
WI-9855	31 A		0.17 C		0.83	0.28	GAAACTGTAGMAAAATCTTTT
WI-9891	39 T		0.44 C		0.56	0.49	ACTGCCTCCTYAGTGAGCCTG
WI-991	37 A		0.63 T		0.37	0.47	TTCTGTACATWCATTATTGTA
WI-9975	126 C		0.88 T		0.12	0.22	GCCTAGAATAYAGTGGTCCC
WI-9983	146 C		0.69 T		0.31	0.43	AGCATTATGAYAGACACAAAG
WI-9986	42 T		0.75 C		0.25	0.38	ACAATTTGAAYGTACCCCAAG
WI-14263	49 T		0.63 C		0.38	0.47	AAAAGGCATATTCAAYTGTCCTACTAATT
WI-14267	28 T		0.94 C		0.06	0.12	ATTAGGAAGGAGCAYTGAAATGGGAAGGGG
WI-14284	55 C		0.94 T		0.06	0.12	TTTAGTCAAAAACAYATGCCATGCGGAA
WI-14288	85 G		0.38 C		0.63	0.47	CTGCTATCCAGATSAAGATTGGTGGAAG
WI-14297	86 A		0.81 T		0.19	0.30	GGTACTTTTTCCAAGWAAAATGTTCTGAAT
WI-14319	83 C		0.19 T		0.81	0.30	AGGCACAAAGCTAAGYACATGCAACAATATA
WI-14323	78 T		0.75 C		0.25	0.38	AAGAATCAAAACATCACTGGACCATGGGAA
WI-14323	86 C		0.94 A		0.06	0.12	AACATCATTCTGGACMATGGGAACCTTGAA
WI-14339	102 T		0.81 G		0.19	0.30	ACAGTACATGATTACKGGTTCCAGAAATC
WI-14372	86 A		0.94 G		0.06	0.12	TCAAATAAATAGGGARTTCTCTTTAAATAAC

WI-14373	95 A	0.94 G	0.06	0.12	CCCTGGACGAAACCARCACAATACATCAT
WI-14379	102 C	0.44 T	0.56	0.49	GGGTTATGTCACACCCYGTCAACCTCAAAAC
WI-14408	60 T	0.69 A	0.31	0.43	CACTATTACAGGCTGWAAGTAACAAATGAG
WI-14482	17 G	0.88 A	0.13	0.22	AGAACCATTAAATAARAATCTGCAAGTTTTC
WI-14492	92 A	0.69 T	0.31	0.43	AAATTACTAAATTAAGTGCTTAAAAGAAAAT
WI-14510	104 A	0.25 T	0.75	0.38	TATGCATAACAAAATWTGCCAGTTTAACCAT
WI-14528	62 T	0.75 G	0.25	0.38	CTGGATGGTATAAATKTTGAATTATAAATTT
WI-14546	95 C	0.81 A	0.19	0.30	ATAGTAGAGGACTCAMCCTGCACGTGCACCT
WI-14580	100 G	0.69 A	0.31	0.43	CCCATCTGTCTTGCAAGGAGGGATCTTGGTC
WI-14631	82 G	0.94 A	0.06	0.12	TCIGICTTCTTTAACRTGCCTGGTCCCTCT
WI-14635	22 G	0.94 A	0.06	0.12	AGATACAGAGCTGTCRTCTTGAAGACCACCA
WI-14651	49 C	0.88 G	0.13	0.22	CTCATTTAAAATTTGTSAAATAAGTCAGAAAA
WI-14666	105 T	0.63 A	0.38	0.47	AGCTAATGTATTAAAWAACCATGAAAAAGAAA
WI-14683	91 A	0.88 T	0.13	0.22	TAGTATCTAAAAACAACAWCAAAAAACACTGG
WI-14712	38 T	0.63 A	0.38	0.47	TCCAAGTACAAATCAWCTCACAATACCATAT
WI-14733	98 G	0.50 A	0.50	0.50	GACAGATATTCTGCAAAATAAATGGCCTGAC
WI-14759	73 T	0.56 C	0.44	0.49	GTTTGACTTGCGGYGTACTCAAATGGGGG
WI-14808	52 T	0.69 A	0.31	0.43	ACCACACTACCCTGTWAAAACTTAACATTG
WI-14816	29 A	0.69 T	0.31	0.43	GAGTCAGCATTTATTWAAAAAACTGGACACGC
WI-14836	28 T	0.94 C	0.06	0.12	AGAGGACAGAGTGTGTTGATTTCGTTT
WI-14856	60 A	0.88 T	0.13	0.22	CGAAAAATACCTAATWTAAAGTTTGTAATAA
WI-14863	61 G	0.94 A	0.06	0.12	AATATTTTGTCTGRAGTTAATAAAGTTAA
WI-14867	46 T	0.56 C	0.44	0.49	CAAGGCTCTCTAACAYGAGTGTCTGCAGCCC
WI-14898	50 A	0.88 C	0.13	0.22	GAAGAGTTGCTCATMAGGTGCCACTAAGGA
WI-14898	79 A	0.88 C	0.13	0.22	GAAAACTTCTCCATMAAGCTGCCTGCTGTG
WI-14907	48 G	0.81 A	0.19	0.30	ACATTGGACTCTGACRATTCCTCTGCAGCA
WI-14911	52 G	0.38 A	0.63	0.47	ATTCAGTTCCTGGTCRAAGGTCTTTTCCTG
WI-14913	88 C	0.88 A	0.13	0.22	ATAGTAGAGGACTCAMCCTGCACGTGCACCT
WI-14914	66 G	0.63 C	0.38	0.47	CAGTTTCTCTAGCAGGAATTTATTGTCCTG
WI-14926	49 T	0.94 C	0.06	0.12	TGGGCACTTAGCGAAYACTTGTGGACCAAA



WI-14930	55 C	0.81 T	0.19	0.30	GAGTCCCTCATGGATGCGGTATTGGTTGGT
WI-14946	47 T	0.94 C	0.06	0.12	CCCCAGACATAACAYCTCTAAATCATCCTC
WI-14948	56 T	0.13 C	0.88	0.22	CTGCTAACTTGTCTAGTCCCACTGATGT
WI-14958	83 A	0.75 G	0.25	0.38	CTTCTTTTCAAGGGRAAAACCCAAATGA
WI-14976	35 C	0.44 T	0.56	0.49	TTGCTTCGTTCAAAGYGCCTTAGAATGGAAGA
WI-14981	31 G	0.38 T	0.63	0.47	GTTATTGGATTTTCTTTTATGCTAAGTATT
WI-14992	80 C	0.25 T	0.75	0.38	TAAATGAAGCTGCAGYAGGAAAGCTGAGCAC
WI-15000	90 G	0.88 A	0.13	0.22	CAGACTGTCTAAGTARTGAAGTTTGTGCAGA
WI-15002	72 T	0.94 A	0.06	0.12	GCCTTCCTGATTTCCWTTTCAGTTTAGGCCTC
WI-15012	59 G	0.56 T	0.44	0.49	TTTCATTGAAGCTTTTCTACTTACTATACTC
WI-15069	81 T	0.94 C	0.06	0.12	ACGCCTAAAAAAAYGTGTGCTTGTGCTGCTG
WI-15100	74 G	0.94 A	0.06	0.12	GACTGGAGTGAGAACRGTTCCACCACCAAG
WI-15116	96 C	0.81 T	0.19	0.30	CCCTAGTTGCAGTAAYGTGTCTATAATAATA
WI-15123	55 C	0.63 T	0.38	0.47	CAGATAAATAGGATGYGTCTGTTGCCCTTA
WI-15152	51 G	0.94 A	0.06	0.12	CTATGTAACACACARTATGCACACCACAGC
WI-15153	40 A	0.81 G	0.19	0.30	TATGTTGGCATTGCARAGACACTGCACCTAT
WI-15182	49 C	0.88 A	0.13	0.22	AACGAGGCAAAATAMTGTGGATTAAACCCA
WI-15198	38 T	0.38 C	0.63	0.47	GCCCTTGGCACTATGYCTACTCTGCCTGACG
WI-15215	84 G	0.44 C	0.56	0.49	TTAGAATCAAATGGGTGACTTTTCCCTCTG
WI-15225	80 C	0.75 T	0.25	0.38	ACCTAGAAAGCAAACYGAGTGATTATGCCA
WI-15239	57 T	0.56 C	0.44	0.49	AATAAACACCATCATYCTGAGTCCACAGAT
WI-15249	34 T	0.81 C	0.19	0.30	ACAAAGTTCTAACTTCTGTTAAATAATCTCT
WI-15260	75 G	0.63 A	0.38	0.47	GAAGCTAATCATGGARGCAAGCTCCCTGGAG
WI-15288	108 C	0.63 G	0.38	0.47	AGGATTCCTCTCTCSTCCAAAGGAAAGAAAG
WI-15295	27 G	0.63 C	0.38	0.47	GAATGTAATTCCTGATSTTTTCTTTGCCAAC
WI-15325	39 T	0.13 C	0.88	0.22	ATGTGGCTGGGAGCYTCACAATCATGGTGG
WI-15347	74 C	0.81 T	0.19	0.30	GAAGAACAATAATTTCAAAGACTTGGGGGA
WI-15353	37 G	0.94 A	0.06	0.12	CAATGTGGTGAAACRTCTTAATTCAGGACA
WI-15361	101 A	0.56 G	0.44	0.49	GAACTCAAGTCATCARTTTAGGCACAAAGG
WI-15389	33 G	0.69 A	0.31	0.43	AGCTTGCTTTTGTCTRTTGGAAAGACTACCA

16

WI-15389	104 G	0.81 A	0.19	0.30	AAACATCTCGAAARAAGTGTGGGAATCAC
WI-15407	92 A	0.56 G	0.44	0.49	AAGGATTAAGTTAARCCACACTACCAAAAG
WI-15488	69 C	0.31 T	0.69	0.43	CAGCCAGATATCAACYGTTACAGAAATGAAA
WI-15625	40 C	0.38 T	0.63	0.47	AAAAGGCATATTCAAYTGTCCTACTAATT
WI-15702	48 G	0.63 C	0.38	0.47	AAGGCTTCAAAAAGSGGGGTAAAGGGTGA
WI-15702	90 C	0.56 T	0.44	0.49	GAGAGAAACTGTAACYCTGTAAACAATACTA
WI-15702	101 T	0.69 C	0.31	0.43	TAAACCTGTAAACAAYACTAATGGGTCTTT
WI-15702	107 T	0.31 C	0.69	0.43	TGTAACAATACTAAAYGGGTCTTTGAACAA
WI-15705	50 A	0.13 G	0.88	0.22	ATTTAGACTGAATCRTTCTAGAGTATTGA
WI-15719	69 A	0.63 C	0.38	0.47	TTTCATCCATTGAGCMAAATTTAAACTCTTG
WI-15729	35 A	0.56 G	0.44	0.49	CCATGTGTAGACTGCRGGCACITTTAGAAAAG
WI-15736	27 G	0.81 T	0.19	0.30	CATTAACTTGCAKATAGCAAAAATAATCA
WI-15747	88 T	0.69 C	0.31	0.43	ACTAATTTAGTGTITTTTAAATATATGAA
WI-15801	24 G	0.81 A	0.19	0.30	CCAAGAAATGGAAGCRCATTTTCATTGGCTT
WI-15801	81 T	0.63 G	0.38	0.47	TAGCTGCAGTAATACKGCATCCCATCCACTC
WI-15809	77 T	0.38 G	0.63	0.47	TCTGTTGTAATGCCKTTTACAAACATTGAA
WI-15843	62 C	0.25 T	0.75	0.38	CCAAGAAAGCCTTCAGYAGAGCAAGTCTGAGC
WI-15868	21 G	0.69 C	0.31	0.43	ATGCAATGAATAAAASGGCAGAAAATTCAGA
WI-15892	123 A	0.94 T	0.06	0.12	AACCAAGAGAAAGGAAGGGAATCAACTCCACA
WI-15937	24 A	0.75 G	0.25	0.38	CTGCTGTATTTAAARACAAAGCGTCTGGATC
WI-15944	24 A	0.88 C	0.13	0.22	AACGTATTTCTCCAMACACCGTAGAAACTT
WI-15953	26 T	0.31 G	0.69	0.43	TGTCCTTCACATCATKTATATTGTATTGCAC
WI-15953	59 C	0.56 T	0.44	0.49	AAACTTTTTTAACCTCYGTCAAAAACAACAAG
WI-15964	99 T	0.88 A	0.13	0.22	CTGTCCCTGGAGGTAWGCAAGAGGGTGGAGA
WI-15986	60 T	0.69 G	0.31	0.43	TGTGGGTTTTTTTTTKTTACATTTTCTTTTA
WI-15987	32 C	0.38 T	0.63	0.47	TTAAAGGGGTCCCAAYGAGGTGGTAGTGCC
WI-15987	80 A	0.88 G	0.13	0.22	ACTAAGAAGATGGTCRTCTATGAACCAAGCT
WI-16002	59 T	0.25 C	0.75	0.38	ATCATGAGAAATTTTCAYGTTAAAAGTCAAAGA
WI-16083	89 C	0.88 T	0.13	0.22	AAACATATCAAGGATYGGGCTGGAATCTTTT
WI-16100	52 A	0.69 G	0.31	0.43	TTTCCTACACTTGACRGTAAATATACTGTTTT

WI-16156	97 A	0.56 C	0.44	0.49	TTAACCCAGAGTCGCMCTCTCTTCAAAATGCA
WI-16163	35 C	0.50 T	0.50	0.50	ATGCAATTGAAATAAYATTGTAAGTTAATGT
WI-16167	58 T	0.88 C	0.13	0.22	TTTCTGATATACATTYCATCTTATTACCAC
WI-1011	70 G	0.86 C	0.14	0.24	AAGTTTTGTCTCCASAGAAGTCATTTTGTA
WI-1172	17 C	0.57 A	0.43	0.49	AACGTGTGGTTAAAAMTAGGCAATTGGTTAA
WI-1172	179 C	0.43 T	0.57	0.49	ATGGCTGATACCAAGYCTGCAGTGAAAAATG
WI-1177	35 G	0.14 C	0.86	0.24	AAAAAATGAAAGAAASAAAGAAAAAAGAGTC
WI-1231	126 T	0.71 C	0.29	0.41	ATTCTCTTCTTTCAYTAATTTCTTTTCACG
WI-1231	141 G	0.71 A	0.29	0.41	TTAATTTTCTTTCACRTTATTCCTCACCT
WI-1319	40 A	0.50 T	0.50	0.50	CATAGTTTATTCTTTWACCATAGGGGTGTGT
WI-1356	123 T	0.79 C	0.21	0.34	CAAGAAAAAAGCCYGTACATGTTGGTAC
WI-472	114 G	0.86 C	0.14	0.24	TATACAACAGAAAAAGSGGCTGGAAAAAGAA
WI-478	46 C	0.64 T	0.36	0.46	TACTCTATTTGTTCYAGCCACCTGTGGCAT
WI-533	29 T	0.36 C	0.64	0.46	AGTACCTTTCTAACTYATAAGATTGTGTAGA
WI-601	74 C	0.07 T	0.93	0.13	AAAGATGGTAGTGAGYGAACACAGAAGAGTTT
WI-601	112 T	0.64 A	0.36	0.46	TCCTAACTGAGTACWCACAAAAACGAGCAGGT
WI-863	107 A	0.64 G	0.36	0.46	TTACAACCTCACCARACTTGGCTTACCGGG
WI-919	36 G	0.64 A	0.36	0.46	TTAATCAACCTAGCCRGCTGTCATGTGGGAT
WI-1736	175 C	0.92 T	0.08	0.15	TCCATCTGTCTTCCAYAGAGATCTAGGGTGT
WI-1754	177 G	0.33 A	0.67	0.44	CTTAAAGAGATAGTCRCCAGAGGCAATTGCA
WI-1775	47 C	0.83 T	0.17	0.28	ATGGTCTTTTCTCTGYTTTACATCATTTGCA
WI-1851	136 G	0.83 A	0.17	0.28	TATTAACATGGTACARACAACCTCAGTTTAA
WI-1949	86 T	0.42 G	0.58	0.49	TGAGATGCTCTGAGTKCAAGGCTGCTGACAT
WI-1949	160 T	0.50 C	0.50	0.50	ATGAATGCCATAATCYCTGTGTTTTTTGTCC
WI-1965	105 G	0.67 C	0.33	0.44	AGGAAGTGTTTAAAGSAGAGAGATGACCCAT
WI-2020	145 C	0.92 A	0.08	0.15	TGGGTCAACTATGATMCCAAAAACAGCAGTGT
WI-2028	176 T	0.17 C	0.83	0.28	GTTCCCTGTCTCATCYTTCTAGGTAATTTGA
WI-2033	183 T	0.25 C	0.75	0.38	AGAACTAATCCCTCAYGGAGAACGTGGAACC
WI-2034	150 T	0.42 C	0.58	0.49	CAGTGCACCAAGGACYGACCTGCACTCTAT
WI-2038	155 C	0.83 T	0.17	0.28	ATTCTATTTTGATAYTGATGTTCTTCTTCAA

WI-2287	24 T	0.92 C	0.08	0.15	TCGTGGTCCCTTTAYAAAGCCCTCTTGATC
WI-2296	81 A	0.50 G	0.50	0.50	ATCTTTGCTCTGACRCCAGTTAGCTGTG
WI-2300	77 G	0.33 T	0.67	0.44	AGAAGCCAGTCATACKTGCTTTAAAAATTGAC
WI-2371	55 G	0.69 T	0.31	0.43	TTCTTCCAGCTTCTKGTGGTGGCTGCTCAAT
WI-2395	122 A	0.69 C	0.31	0.43	AAAATTACTATCCAAMCTGAATTCAGAATAA
WI-2437	128 G	0.06 A	0.94	0.12	CCAAAAATCCCAATRCTCTAAATAGATGGA
WI-2437	179 G	0.94 A	0.06	0.12	CAAGAGGCAATCGACRAACATCACAGTGGC
WI-2437	192 G	0.94 A	0.06	0.12	ACGAACATCACAGTGRGCTGTGGTGCCAAAGG
WI-2440	71 G	0.88 A	0.13	0.22	ATTTAAATTTAGTTGRGTGAGACCAATAGCA
WI-2572	61 C	0.94 T	0.06	0.12	AACACTTCTCCACAYACAAAGTTAACACTT
WI-2616	125 T	0.13 C	0.88	0.22	CAAGAAATTGATCCTAYACTGGGACTACAGCC
WI-2625	98 G	0.00	0.00	0.00	AAGGCTATTAGGA CAAATTGATGATACT
WI-2716	23 T	0.88 C	0.13	0.22	ATCCAGAAAAACAGCYGAATGACAAACAAGAG
WI-2886	46 C	0.81 A	0.19	0.30	GTCTGGGGAGAAGAMAACGAGATAAAGCAT
WI-2906	50 A	0.25 C	0.75	0.38	CTTCATCTTGCTGGMACTTTGCCTGGAATG
WI-2906	77 T	0.31 A	0.69	0.43	AATGCTCTTCCCTCWGAGCTTTGCTTGGCT
WI-2924	54 G	0.75 A	0.25	0.38	GCTTCTCTTATAGGRACCCCTGTGATTACAC
WI-2939	72 G	0.63 T	0.38	0.47	TGTCTCAGTGCCTTTKCAAGACCTTCCCTCA
WI-3000	62 G	0.38 A	0.63	0.47	AAACACAGAGACCCCTGAGTCTTAGTCAAT
WI-3167	37 T	0.88 A	0.13	0.22	AGATCTATTAGATTTCWCACCCCATCTCAAAAC
WI-3203	99 G	0.63 A	0.38	0.47	TATGCCGACAGACGAGRCCACACAAGGCAATA
WI-3208	140 G	0.69 A	0.31	0.43	GTGGCAGATAAAGARCCAAGCCCTAGTTTG
WI-3275	157 C	0.94 G	0.06	0.12	CAGAACTATTTCTCASTAAGAATCTTAAGTT
WI-3402	55 G	0.50 A	0.50	0.50	TTGATTTCCCTTACATRCAAATGCTCCTTTT
WI-3416	33 C	0.69 T	0.31	0.43	TAGCATTCAGAAGTCYCTCTTAGAGGTAGTT
WI-3453	70 C	0.19 T	0.81	0.30	GGCCCATCAGAGAAATYGAAGTCATGGGAAA
WI-3473	101 A	0.88 G	0.13	0.22	TTTTAGCCCTAGGGARTAGAAAATGTTGGTG
WI-3474	90 A	0.38 G	0.63	0.47	CCCTAAATTTAGCACRGATTTTAAATGAGGT
WI-3474	109 G	0.94 A	0.06	0.12	TTTTAATGAGGTGGTGGTGGGAGAAAATTGAT
WI-3502	79 C	0.56 T	0.44	0.49	GGTTCTGGATGCTCTYTGAGGACAGGGTCAC

WI-3600	78 T	0.88 G	0.13	0.22	CCCTGATAGTTCTGKGAGCCACCTAAACTC
WI-3600	146 G	0.56 C	0.44	0.49	TGGATATAAACATCTSATGGAAAGGCTGCACT
WI-3687	67 A	0.94 C	0.06	0.12	AATATGACATAAAATMAAAACTACTATAGT
WI-3735	72 T	0.63 C	0.38	0.47	TATCAATGAAAAACYACACCGGTTCAATGA
WI-3746	116 G	0.94 A	0.06	0.12	CATCTGTCTCTGCRGCCCCAGGATAAAGC
WI-3867	49 T	0.69 C	0.31	0.43	TAGTCTTCTGACAAAYCGGATGTACCTAGTA
WI-3898	25 A	0.71 C	0.29	0.41	TGTCTTTAGAAAGCAGMGAGAGGACACCGAC
WI-3901	114 A	0.07 G	0.93	0.13	TCACCTGACAAAGTGGRATCATGTGCTACAC
WI-3914	99 C	0.71 T	0.29	0.41	CTCAAGACTCACAGCYACCATCCTTCATTGC
WI-4019	33 G	0.36 A	0.64	0.46	CGTCCTATGAATCATRCATTTGTTCTGTTA
WI-4091	84 A	0.71 T	0.29	0.41	CTTAGTCATTGCATGWTGTATAACAATATTG
WI-4160	117 A	0.86 G	0.14	0.24	ACAATATCAACAGAAARGGCTATATTAGAAAA
WI-4168	32 A	0.86 G	0.14	0.24	AAATTGATACAAACARTCTGAAAACTCTGTTT
WI-4177	68 T	0.64 C	0.36	0.46	TACCTATTATATTTAYCATCATGATTGCTG
WI-4199	51 A	0.43 C	0.57	0.49	AGTCAATATAAAAAAMCACACATATTGTTAT
WI-4250	94 G	0.36 T	0.64	0.46	GTCTTGTGAAACAGGKGTTGGGAAGGATCCTG
WI-4250	117 A	0.57 G	0.43	0.49	GGATCCTGTAAAAAGGRTAAATATTGTTTCC
WI-4255	68 G	0.86 C	0.14	0.24	GCTCCCCCATCACCTSCCTTACACAACCTGA
WI-4256	57 C	0.93 T	0.07	0.13	AGAGGCAAAATCTGGYCTCACCATTGGAAAA
WI-4325	58 C	0.93 T	0.07	0.13	GTACATGGGCAGGACYGGAATGGGATGCTA
WI-4325	71 C	0.57 T	0.43	0.49	ACCGGAAATGGGATGYTACTATAGATAATCT
WI-4347	158 A	0.07 G	0.93	0.13	TATCTGTTCAGGCCCRGAATCGTCACGGCTC
WI-4360	93 C	0.63 T	0.38	0.47	GTATTTTCCAAATAAYAAAAATGCCCTCTGAAA
WI-4448	112 T	0.63 G	0.38	0.47	AGATGGGGTATATAKAAAGAACCATGTAAA
WI-4456	49 C	0.69 T	0.31	0.43	GAAAATTATAGTTCYCAAGTTTCATGCATAA
WI-4461	49 A	0.50 G	0.50	0.50	TAAAAATTATCCTTCCTCCRTGAAATGGTGAAAG
WI-4465	41 A	0.75 G	0.25	0.38	AGACAACACGAAAGTRTATAAAGAAAAACAGT
WI-4465	75 G	0.75 A	0.25	0.38	TAATCTTTCACCTTTRTATTTCTCTCTTACC
WI-4529	64 T	0.44 C	0.56	0.49	ATCATCTGAAGATGYGAGTCTCTCTTTAT
WI-4540	110 A	0.88 G	0.13	0.22	CACCATGTGGCATCCRTGCATGGCTGCATTG

WI-4596	69 T	0.25 A	0.75	0.38	AGAAAGCACTGTGACWCATTATTAGGCCCAT
WI-4606	61 A	0.56 G	0.44	0.49	AGAAAATTATGCCTARCCAAGTAGACAACTT
WI-4649	50 C	0.44 T	0.56	0.49	CATTCTTTCCGAATGYGATGATTTCTTGTA
WI-4650	148 A	0.13 G	0.88	0.22	TCTTATATGCTTTTRCCAAATCCAGTTAA
WI-4677	82 T	0.69 C	0.31	0.43	GAGTTGAAATAAATGYAAGTTGAATAATGAC
WI-4698	135 C	0.94 G	0.06	0.12	GGAAGAAAACCTTCAASTTCGAGAAGGCCTAG
WI-4722	88 G	0.81 A	0.19	0.30	TATGGAACACACACRCAACTGAATGCAGAT
WI-4745	131 T	0.75 C	0.25	0.38	TACTTTCTACTCTGAYAGGCAGACTTATATG
WI-4782	113 C	0.63 T	0.38	0.47	ATAACTAGAAAATGCGYGAACAGAAAAATAAC
WI-4788	65 A	0.75 G	0.25	0.38	ATCTTGCTAAGTTCCTGAAAAAAAATTATG
WI-4818	43 A	0.38 G	0.63	0.47	GACTAGGTATGTCCRCACATGAATAAACAA
WI-4818	121 G	0.56 T	0.44	0.49	TAAATGGGCCCTGTGCTCTGCGCATACATAT
WI-4888	56 G	0.81 A	0.19	0.30	GAAAAGATAACAAGARATGAATAAATGAGGT
WI-4897	93 A	0.94 G	0.06	0.12	AAAATAAGCGCTTGGRGATAAACACATCTTC
WI-5163	24 C	0.38 T	0.63	0.47	CACTGGTCTGCTGTGCTGCTGCTGCTGCTGT
WI-5204	54 C	0.94 T	0.06	0.12	TGGGTTTTGAAGAAATATRGCGAAAACTATCTT
WI-5215	70 A	0.81 G	0.19	0.30	CAGACTCAAAAATATRGCGAAAACTATCTT
WI-5248	38 G	0.38 C	0.63	0.47	GCTGCTAGTTGTTASAGCAACCCAGAAAA
WI-5248	99 C	0.31 T	0.69	0.43	TATTGACCGTACTTGCTGCTGCTGCTGCTGCT
WI-5252	119 A	0.94 C	0.06	0.12	GTGAATCATTTGCTTTMTACCATGTACATATT
WI-5257	77 C	0.75 A	0.25	0.38	CATGAAGCAAAGAGGMCCTTTCATCTGCCCT
WI-5300	38 T	0.88 C	0.13	0.22	GAGACCATTTCATTCYTTTTTGGATTATGAA
WI-5317	139 T	0.56 C	0.44	0.49	CTGGTAGCAGGTATAYGGACTCATTTCTCT
WI-5328	44 A	0.94 G	0.06	0.12	ACACTGAAAGACAGRAAAAAAAGAAATATT
WI-5345	29 G	0.94 A	0.06	0.12	AGTTTTAAAAATCTRCCTGCTATGGTTTGC
WI-5370	143 T	0.75 C	0.25	0.38	TAACTAATAAACAAAYTTTGAAATTTCTCTGT
WI-5406	42 A	0.94 G	0.06	0.12	AGACTCTTCCAGAAAGGCCACTTCCACAGAT
WI-5406	118 C	0.63 A	0.38	0.47	TGTCAAGGTGAGAAAMCCTATGAGCCACAC
WI-5406	120 C	0.81 T	0.19	0.30	TCAAGGTGAGAAACCTATGAGCCACACTT
WI-5415	54 T	0.75 A	0.25	0.38	TTCACTTTTCAGTTTWTAGATCGGATCATGA

WI-5437	41 C	0.19 T	0.81	0.30	AGAAAAATCCAAGAGYCTTAAACCATATTTT
WI-5481	29 G	0.44 A	0.56	0.49	TTAGTTGATGAATTTTAAATTTTACAGTATCT
WI-5481	131 A	0.31 G	0.69	0.43	TTTATGCTGCAGTCGRAATACTTGGAGCCTG
WI-5492	38 T	0.94 C	0.06	0.12	CTTGTTAAAGTCCCAAYCAAAAGAAAGGATCCC
WI-5546	40 C	0.81 T	0.19	0.30	TGAAAAAAGGGAAAYACCCATGTTTGCTAA
WI-5552	97 C	0.69 T	0.31	0.43	CAGCCTTTTAGAGTYCCTGGCAATTTGTG
WI-5573	58 C	0.75 T	0.25	0.38	ATAAGGAGGTGGGAYGACACATTACTCTCC
WI-5612	44 T	0.94 A	0.06	0.12	TAAATCATTTCAACAWCACAAATATCTTAT
WI-5612	125 A	0.81 T	0.19	0.30	AGCATCGTGTCAATTCWCAGTGTTTTAGGTTT
WI-5636	26 A	0.25 C	0.75	0.38	TTTATCCGCAATAAAMTCCCAAAGTCTCTCG
WI-5752	36 A	0.88 T	0.13	0.22	CTCAGTTTTCATCWTTTTTTTTCATAATTTA
WI-5791	44 C	0.94 G	0.06	0.12	TATTTGGATAAGTTTACAAAGATGAGAACA
WI-5791	76 G	0.88 A	0.13	0.22	GTCCTAGAACCTCAGRATCGAAAGGAAGTTC
WI-5798	48 G	0.88 C	0.13	0.22	CCTTGTTTCTTTTGSATTGAAAAATACTGG
WI-5836	161 C	0.94 T	0.06	0.12	ACATGATTCAATGATYCCATTTGAAAAATTA
WI-5850	92 C	0.94 T	0.06	0.12	GGCTTCCTCTATGCAYGCGTCTATCTTCTAT
WI-5850	134 G	0.88 A	0.13	0.22	TCCAATGTCCCATTCRTTTTGCCATTTCCTG
WI-5874	76 T	0.63 G	0.38	0.47	TACAGAAAAAAATTKTACATATCAAAATGAC
WI-5944	52 A	0.69 G	0.31	0.43	ACCATGGGAATCTTGRTGCAAGTTAGATCCC
WI-5989	29 G	0.44 A	0.56	0.49	CAAAGGTCACAGGCARCGTACATACGGTTCT
WI-6053	24 A	0.94 G	0.06	0.12	GTGCTAAGAACAAACRTCTTCATGTCCAAC
WI-6141	80 T	0.88 C	0.13	0.22	TCTACAAGGIACCTAYCACGTGTTCTGGGTT
WI-6192	91 A	0.50 G	0.50	0.50	GGATTTAATTTGGATRAATTTTAATACTAGC
WI-6194	105 T	0.88 A	0.13	0.22	ATGATAATAAAGAAAWATGCAGACTACACTC
WI-6217	131 C	0.94 T	0.06	0.12	AGCAGCTCATTCAAGYGGCCCAACCATGGCCC
WI-6272	86 C	0.31 T	0.69	0.43	AGGAAAAACTTTAATYTTCTTTGTCTCTCC
WI-6303	96 G	0.63 A	0.38	0.47	AGAAGCTCTGTCTGCRCTGCAAAAGCCATGGC
WI-6375	28 A	0.88 G	0.13	0.22	TATGAAATCAATAGRTATCTTTTACAAAAA
WI-6409	73 A	0.94 T	0.06	0.12	CAAAATCAATTACAACWATGTGCTTATCAGCT
WI-6409	112 T	0.69 A	0.31	0.43	ACCCCTATATTTTAAWGCAACTGACAGTTTTT

WI-6450	45 T	0.63 G	0.38	0.47	CTATATCTTGTCACAKAGAAGTACCACACAT
WI-6461	88 C	0.94 T	0.06	0.12	TTCTATAAAACAACAYAGGAACGAGGCTCA
WI-6523	165 G	0.69 T	0.31	0.43	TAGAGACTGAAGCTGKTATCAACCTTCCCTA
WI-6558	42 G	0.94 C	0.06	0.12	TTTATTAAGGACATTSTGTAATGTTTCCACT
WI-6558	68 C	0.56 T	0.44	0.49	CCACTTTGTTTTAAAYAAATTACAAACATGTG
WI-6629	75 T	0.81 C	0.19	0.30	ATAAAAGTTGTCATAYAGCAATGGATGCTGT
WI-6686	151 A	0.44 G	0.56	0.49	CCAAAACAAGAAATRAACATTGGAATAGTC
WI-6690	28 T	0.38 C	0.63	0.47	CATTATTAAGGAGAGYACTAGGAAAAACTAC
WI-6690	106 C	0.38 T	0.63	0.47	CTCTGGAGCCACAGCYGGCTAATACACTGCA
WI-6761	32 C	0.38 A	0.63	0.47	ACAGCTGCAGAAATGGMCTTCTTCCCTCCAG
WI-6770	53 A	0.13 G	0.88	0.22	CCCCAAAACATCACARAAATTATTCATACAT
WI-6889	139 T	0.88 C	0.13	0.22	ATGCAGTTAAATTCYAGAAATAATTAAAGC
WI-7059	43 C	0.88 G	0.13	0.22	AGGCACCCAGCCATCSTGACCCAGCGAGGAG
WI-7254	37 A	0.75 G	0.25	0.38	TGAGAGAGGAGCCACRGTCCCTAATGACACC
WI-7286	65 T	0.44 C	0.56	0.49	AGCTTAACTGACAGAYGTTAAAGCTTCTGG
WI-7374	182 T	0.94 A	0.06	0.12	TTGAAGAAATATATTGWCAGAAACACAGGCT
WI-7386	104 T	0.94 A	0.06	0.12	TGTAACAATTTGTTAWGTGTTTAGAATCAGA
WI-7423	107 T	0.44 C	0.56	0.49	GCTGGGCTGTGTTCCYCGGGCTCTTCTGGAC
WI-7424	131 T	0.44 A	0.56	0.49	GAGAGGAAAGAAAAAWACAACCTTTCATTCTT
WI-7466	80 T	0.75 C	0.25	0.38	GGCTATGAAATAGTCYATTCAAGTGAAC TAGT
WI-7466	141 G	0.50 A	0.50	0.50	CAGTCTTTGTCCTGGRAATATCTCACAAAAT
WI-7593	46 G	0.06 A	0.94	0.12	AGGATGAAAGGAGAGRAATGAGATCAGTTTT
WI-7753	52 A	0.19 G	0.81	0.30	CCGAGAAGAACAGATRATCCCTGTATTCAA
WI-7836	120 T	0.56 C	0.44	0.49	ACAATGCAACGTTCCYGAATTTCTAATCTTGG
WI-7848	142 A	0.44 G	0.56	0.49	TTTTAAAAACCGTCTCRTGTCTGAATAGCTTT
WI-7858	91 T	0.44 G	0.56	0.49	CGTGAATTTTTAAATKTATAGATGTAAACTT
WI-8172	136 C	0.63 G	0.38	0.47	TGTTTTCTTGACATASAGTACCTTTACAGGT
WI-8183	56 G	0.81 A	0.19	0.30	AACAAATTTCTGTGCRGAGGTTTGATTCA
WI-8377	63 A	0.94 G	0.06	0.12	CCCAGGCCCTTCCRTTATATCCAGGTATG
WI-8540	73 T	0.88 C	0.13	0.22	CCTGCATTGGCTTAYGTGCTGAAAGAA



WI-8550	32 G	0.50 A	0.50	0.50	0.50	TCAATGCAACAAGTARAATTTGTAAACTCAA
WI-8655	29 A	0.44 G	0.56	0.56	0.49	AATAGGAAACCAGAGRGGGAGCCCCAGGTGG
WI-8712	44 G	0.25 A	0.75	0.75	0.38	GAAGAGGTAGTGGAGRGAGATGGTCAGGCTT
WI-8827	22 C	0.19 T	0.81	0.81	0.30	CCTGGGAGACTATGGYAGTGAACACTAAAT
WI-8833	51 A	0.88 G	0.13	0.13	0.22	CCATGCCATTCTCTGRTGCCCTATAATGTG
WI-8850	21 A	0.50 G	0.50	0.50	0.50	CTTAACCTTTGGCTRCCTGCCTGGCTGTTT
WI-8853	79 C	0.50 T	0.50	0.50	0.50	CGGCCATTGAGGATAYATGGAAGGCTCAGGA
WI-8865	42 T	0.31 C	0.69	0.69	0.43	TGAGGAAGACAGTCAYGGTCGAACAACAAC
WI-8865	52 A	0.56 G	0.44	0.44	0.49	AGTCATGGTCGAACARACAACATGCTTCGGA
WI-8895	32 A	0.94 C	0.06	0.06	0.12	ACCAACCAACAGAAATMCTCCCGTCTTTGAA
WI-8974	34 C	0.38 T	0.63	0.63	0.47	GCCCTCAAGAACTCAYGCCAGCTCAGCCCTA
WI-8997	41 G	0.81 A	0.19	0.19	0.30	GCCCACTTGCTCCCRCTGAGCACTGCGTACA
WI-9005	26 C	0.81 T	0.19	0.19	0.30	TTTGCTGGGAATCTYGTITTTCTTCTTAAG
WI-9014	18 C	0.88 T	0.13	0.13	0.22	TGTTCCCATGCTGACYTGTGTTCTCCTCCCA
WI-9014	44 C	0.31 T	0.69	0.69	0.43	CCCCAGTCATCTTCTYGTGTTCCAGAGAGGTG
WI-9014	93 T	0.63 C	0.38	0.38	0.47	TCTGTCTCAACTTAYGTGCACTGAGCTGCA
WI-9015	48 C	0.00 T	1.00	1.00	0.00	AATTGGCTGGATTGYGCTTTGGTTAATACA
WI-9063	53 A	0.44 C	0.56	0.56	0.49	AAAGACACCAATTTATMTACCCAAGGCAGAA
WI-9064	29 A	0.44 G	0.56	0.56	0.49	AAACATAATTGATTCTATCTGCGAGACTTA
WI-9074	38 A	0.63 G	0.38	0.38	0.47	TTTGCTCTAAAGAARAAGGAAC TAGGTCAA
WI-9161	61 C	0.50 T	0.50	0.50	0.50	TAAGCATTGCCTGGCYTTCCTGTCTAGTCTC
WI-9171	62 G	0.94 A	0.06	0.06	0.12	TAGAGATAATAATCARITCTTTACAACCGAT
WI-9174	47 T	0.56 C	0.44	0.44	0.49	CCATTCTCTATTTAYCAGTCTCTGCTCTATA
WI-9186	76 G	0.63 A	0.38	0.38	0.47	CCACTTCTCCCGCARACCTAGGTGAGACTT
WI-9193	94 G	0.69 A	0.31	0.31	0.43	GTCTGCCTTAAGCARTACCCCTACCACA
WI-9231	32 G	0.75 C	0.25	0.25	0.38	GGTCCCCAGATTGASGTCTGAGTGTGGCA
WI-9274	25 C	0.44 T	0.56	0.56	0.49	GACTTCACCTTTGGTGYCAATGGACAGAAAAAT
WI-9281	68 G	0.94 A	0.06	0.06	0.12	CTTGCTGGCTACTGGRTGTTAGTTTGCAGTC
WI-9304	70 G	0.25 A	0.75	0.75	0.38	ATGATCACCAGCTGARAATATTGTTTACAA
WI-9343	78 C	0.81 T	0.19	0.19	0.30	CAACATCCTCTCTGCCAYACACAACAACGTA

WI-9357	75 A	0.94 G	0.06	0.12	GTTATTATGCTCTTARTGATTACAGACTGA
WI-9360	79 T	0.69 C	0.31	0.43	TCTGCTTTAACTTGGYATTCCTCTAATTGTG
WI-9413	112 G	0.38 C	0.63	0.47	CTGCTATTCCAGATSAAGATTGGTGAAG
WI-9557	74 C	0.88 T	0.13	0.22	GCCAGCTACAGCCTYGGTGCATCTTAACCC
WI-9720	47 A	0.00 G	1.00	0.00	AAATACCCCTCTCTRATAATTTAAGTAACC
WI-9720	55 A	0.00 G	1.00	0.00	CTTCTAAATAATTTTRAGTAACCAAAATATT
WI-10019	139 A	0.88 T	0.13	0.22	TATGTAGCAAATCTAWTCCCCTAAGCACAGT
WI-10020	39 T	0.56 C	0.44	0.49	GTATTAAATAAATAYGTTAACTGGCTCTGA
WI-10020	122 T	0.88 A	0.13	0.22	AAATCATGACTTTTWWAAAAATACCAGACTA
WI-10064	54 C	0.81 A	0.19	0.30	CAGGATCAGGGAAGGMATTATAATAAATATA
WI-10064	170 C	0.81 T	0.19	0.30	TGATTGTTTTACATGYGAAATCTGGCCTCAG
WI-10289	29 T	0.31 C	0.69	0.43	GTCCCCAAACTCTTAYTTAATTCCTCAAT
WI-10316	104 T	0.44 C	0.56	0.49	ACCTCTATTCTCTTAYTAAACTTTTGATAC
WI-10368	31 C	0.50 T	0.50	0.50	CAACCAGGTCTGTCTACCCCTCTTAGAG
WI-10391	32 A	0.88 G	0.13	0.22	CAGGTATGACTCCCACTCAACTCTTGACTC
WI-9748	74 C	0.94 G	0.06	0.12	TACCCCTTGTCAATSTCAGACCAAGTACAT
WI-9763	21 G	0.75 A	0.25	0.38	AACTCTGCGGTGTGRAGAAAGGACAGTTAT
WI-9897	83 A	0.63 T	0.38	0.47	ATTTATCTAGCCTGTWCAAGTCATCCAGTGA
WI-9897	84 C	0.88 T	0.13	0.22	TTTATCTAGCCTGTAYAAAGTCATCCAGTGAG
WI-9935	42 C	0.56 T	0.44	0.49	TAATAACGTGTGCAAYACCTCACCAGAACTG
WI-9935	115 C	0.38 A	0.63	0.47	GGGGAGTTCAGACAMAGCCCAAGAAAAGCCT
WI-9943	91 T	0.81 C	0.19	0.30	TTTATATCCATCTTCYATTTTAAATTTCTAC
WI-10567	60 T	0.13 C	0.88	0.22	AAATATTATCTTTTTCATATTTTCCAATT
WI-10567	82 A	0.94 C	0.06	0.12	TTTCCAAATTAATMCTAGAAATTTTACCACCA
WI-10567	146 A	0.13 C	0.88	0.22	GTCCTTAATAGCAAMAGCTACTGGAAGCGG
WI-10686	133 C	0.81 T	0.19	0.30	TGCCCTGTCCAAGGYGTGTCTACACATGA
WI-10694	144 A	0.75 G	0.25	0.38	GCCTTATGAGTTTCRTTCTCCTCTTACAA
WI-10719	115 T	0.56 C	0.44	0.49	TCAAGGCCATCTAGYGGCTGCTGGCAGTGC
WI-10721	40 A	0.38 G	0.63	0.47	CTCTGCTACTGCCARATGAGATTTATTTAT
WI-10732	80 C	0.63 A	0.38	0.47	CTTCATTGGTCACTMTTAAAGTCTGTAT

WI-10775	39 C		0.75 T	0.25	0.38	TAATTCATTACACTCYACATCATATTTCTT
WI-10778	62 A		0.13 G	0.88	0.22	GAGGAACATTTACAGRGTCCTCTGTATGT
WI-10789	21 C		0.50 T	0.50	0.50	ACACTGCTCTAGACCYTCCAGGGTCCCTCA
WI-10810	58 C		0.50 T	0.50	0.50	TCATGGCAGGAATTTCATTTCTGTGTTCT
WI-10828	23 T		0.94 C	0.06	0.12	CAGAACTACTGGCAYAGGGTTCTTAAAC
WI-10832	91 G		0.75 C	0.25	0.38	ATCTGCAGGCTCTCCSTTTCTAAGTCACCTG
WI-10834	96 C		0.44 T	0.56	0.49	CAAAAGTGTGTTAATYCTTAATACCAATTT
WI-11027	90 T		0.44 A	0.56	0.49	TACGCTTTTAAAAAATAAAAAATACTGTA
WI-11049	95 C		0.06 T	0.94	0.12	TGTTTCAACTAAGGAYAGACTTCAGAAGGCA
WI-11070	110 G		0.38 T	0.63	0.47	TCAGCCAGCTATCTTKGGTGCAGAGAGGTAC
WI-11070	135 C		0.75 T	0.25	0.38	AGGTACTCCAAGTACYGTGGGGTCTGATG
WI-11076	106 T		0.50 C	0.50	0.50	AAGGGGAGCAGGCAYGTCACATACCCAGAG
WI-11076	142 G		0.81 A	0.19	0.30	GAGAGAGAAAGAGAGRAAGTGCCACACATTT
WI-11153	33 C		0.69 A	0.31	0.43	CTCACCTAAATTATGMGTGATTAAAAATATAC
WI-11153	84 C		0.69 G	0.31	0.43	GCTTTAAGTACTTTASGAAGACCTTGACTGT
WI-11163	58 C		0.56 T	0.44	0.49	ATGACCAAAATGAGAYAAATTTGTTAAAAAA
WI-11169	95 A		0.75 G	0.25	0.38	AAAAAATTAAGCCTRAAGTAGTGCTTTTAA
WI-11169	154 T		0.81 G	0.19	0.30	AAAAAAGAGCAGACAKTTTATCATGTGTTCT
WI-11175	77 T		0.81 A	0.19	0.30	TTTCTGCTCAAAGAGWTTTTTTTAAAGTTATC
WI-11204	80 T		0.69 A	0.31	0.43	TGAAAAGAAAAAAGCTTWCACCTTTTATTTAA
WI-11204	88 T		0.94 C	0.06	0.12	AAACTTTCACCTTTTATTTTAAAGTAACAT
WI-11206	127 A		0.81 T	0.19	0.30	CTGTATGTACAACTCWCACCAACCATTAGGATT
WI-11215	68 C		0.94 T	0.06	0.12	CAGATTTATTTTAGTYATTTTCTATAAT
WI-11219	18 G		0.56 A	0.44	0.49	AAAAATGCATTAGAARAATTGGAGGATAAAA
WI-11219	89 G		0.81 A	0.19	0.30	AGATGAAAAATAGGARAGAAAGGTAGAAAA
WI-11222	25 C		0.75 T	0.25	0.38	GAATCATTTACACTAYCGAAATCAGCAAATG
WI-11222	136 G		0.88 A	0.13	0.22	TACCACTGCGGCTGGRTCACAACTGGGTAC
WI-11226	165 A		0.94 C	0.06	0.12	TTTGACTATGAACAMGACATAGTTGCTAAG
WI-11276	41 A		0.44 G	0.56	0.49	CAGCCAGGAGCAGACRCACCGGCTCCTCAGT
WI-11282	42 C		0.81 G	0.19	0.30	CAGAGAGCAAGGGAASCACACAAAAATTTACA

WI-11295	37 A	0.56 G	0.44	0.49	AAATATAATTGCTRTAGAGTTCACAGATG
WI-11305	87 C	0.81 T	0.19	0.30	CACAGCATCACACCAYAGGGCCACGGGAGG
WI-11321	67 A	0.56 G	0.44	0.49	AATAAATTTTTTAAAGGTTTAGCTATTC
WI-11324	40 C	0.56 G	0.44	0.49	AAATCATGTGCCCCASAGAGCCCAAGCTT
WI-11352	69 T	0.75 C	0.25	0.38	GCACATAGTGGAAGYGCTAAGTGTCTACG
WI-11352	104 T	0.75 C	0.25	0.38	GTCAGATCATATCCAYAGAAAAACAGCTCTC
WI-11371	84 C	0.56 T	0.44	0.49	GAGATTCTGATTTCAGYGTGCTCAGGCGGGC
WI-11385	75 T	0.44 C	0.56	0.49	TAAAGTCTCTTCAGYAGGAAAAAAGCTACA
WI-11388	88 C	0.25 A	0.75	0.38	CACGTAACTAAGTTCMTATATTTTAACTTG
WI-11392	55 T	0.31 G	0.69	0.43	AACCTTAATAAATACTCTTTTACAAAAACAC
WI-11396	52 A	0.50 T	0.50	0.50	TTGAAATGGTTTTWTGATGGGTGAATATGA
WI-11441	100 C	0.50 A	0.50	0.50	TCCCCACCAACCAGCMCAAAATAAGGCCCTGG
WI-11466	26 C	0.69 T	0.31	0.43	CCATTATTTTGCAGYCTTCAGTCCAAAAAA
WI-11537	119 C	0.88 G	0.13	0.22	TCCTACTCTGACCATSATAATCATTCTTTT
WI-11549	102 T	0.44 G	0.56	0.49	TCCTTTAAATATCTGKGGGATTTGTACAGA
WI-11585	79 T	0.63 C	0.38	0.47	TTTGCAAAAAACAAAAYGGAAGTATCAGTGAA
WI-11604	68 G	0.94 C	0.06	0.12	CAGTTACCAGCATTTTSAGAACTAGGGACTTT
WI-11614	60 A	0.75 G	0.25	0.38	AGACTCAGCTGCTTGRGGCATGTTCCACCC
WI-11614	108 C	0.75 A	0.25	0.38	ACTGTGAAACTGCAAMATATTAAGTATTCGT
WI-11626	39 G	0.50 A	0.50	0.50	GGAACATGAAGGTAGRGATAAGGTACAGGA
WI-11626	83 T	0.38 C	0.63	0.47	TATTTTAAATAAAAYTACTTAATAAAGA
WI-11627	23 T	0.69 C	0.31	0.43	CCTTCCATTGTCTCTCYCTTGAGATGGGTTGC
WI-11636	61 A	0.88 G	0.13	0.22	AGATCTGCTTATCCCTRTATATCCACATAACT
WI-11654	37 G	0.75 C	0.25	0.38	ACTATTCAGCAACTGSAAACTGTCCTGGGAG
WI-11656	28 G	0.25 A	0.75	0.38	TAGAAGGAACTGCAARCTTACTTGAGGACA
WI-11680	55 T	0.94 C	0.06	0.12	TGATTCTCCCTTTTGTGCATAAAGGCTGG
WI-11696	47 T	0.88 C	0.13	0.22	CACAGCAGGGGACAGYAAGGTGGCTTCTCT
WI-11702	69 C	0.81 T	0.19	0.30	AAATAACCACAGCAGYTTTCAGTATAAATTG
WI-11706	60 C	0.50 T	0.50	0.50	GTACAAATTTATTTGCGYGGCTGGAAATTTTC
WI-11709	105 T	0.44 A	0.56	0.49	CTTGCTTCAGTTTGCWGTCCCGTAAAAATTAG

WI-11710	103 C	0.50 A	0.50	0.50	AGCTCAGTCTTCACMCTCCTCCTCCTCCA
WI-11715	49 A	0.75 C	0.25	0.38	TGTAACACAGACAAAMTGCAATACAACTGTG
WI-11715	123 C	0.63 T	0.38	0.47	GGCTGGCTGCAGCTTYAGCCACAGGATGGGG
WI-11727	43 G	0.38 C	0.63	0.47	AAACAACTATCAACASCTGCAACACAAACCA
WI-11728	16 C	0.50 G	0.50	0.50	TTTATTTATCAAACTSCAATTCATTTCACA
WI-11758	61 A	0.88 G	0.13	0.22	TGTGGTTTTCGCCTGRTAGACCACAGGGCCA
WI-11773	93 T	0.06 C	0.94	0.12	CCTTTTTTTTCCCCCYGTGATTGTTAATTAG
WI-11790	28 A	0.81 G	0.19	0.30	TTACCAACCTCTGTGTCCTTAGCCTCGCCTA
WI-11806	60 T	0.88 G	0.13	0.22	AGAGTGGGCAGTTCAGTTTTATTAGTATAT
WI-11879	61 C	0.81 A	0.19	0.30	GTATTTAGTATACAGMAGTGATTTTCTCTCT
WI-11906	52 A	0.69 G	0.31	0.43	AGAAAGAACTCTGAATRTGAGGGAACCTGCAGA
WI-11909	78 A	0.38 G	0.63	0.47	TGTTGGTGGTCAAGRCTATTTCAGAAAATCT
WI-11946	31 C	0.94 A	0.06	0.12	CTTTGCTCTGGAGACMCCAGCTAGTCTAAGA
WI-11965	65 T	0.56 G	0.44	0.49	CTCTGGTTTATTTAAKATCAACATTCAACCAC
WI-12002	30 C	0.13 G	0.88	0.22	GAATCCAGGACACAASAAAGAAAAACACCCAA
WI-12002	68 G	0.13 A	0.88	0.22	ATGGAGACAGAAGACRAGACACAACCTCCTCC
WI-12002	89 T	0.56 C	0.44	0.49	CAACTCCTCCCCCACYGCCTCCTGCTCTAG
WI-12018	31 A	0.56 T	0.44	0.49	AGCCAGCTCTGACTTWTCTCTCTGTTTCTGTC
WI-12020	121 T	0.94 C	0.06	0.12	GAATACATGACCAATTCTCTTTTAGCAGGTT
WI-12075	103 G	0.50 A	0.50	0.50	GGGCACGGGGGAGGCRGAAGGAAGAGAAAGA
WI-12086	72 C	0.81 T	0.19	0.30	GGAAAACTTGGATTTCYCAAGACCCGAAGAC
WI-12108	40 C	0.31 T	0.69	0.43	TTAAACTCAAATATCYGAAATACTTTTCATTA
WI-12159	28 C	0.50 T	0.50	0.50	ACACCGTGCAATGCAAAAGTGCACTGAGGA
WI-12169	121 G	0.81 C	0.19	0.30	TATTTCTTTTGCTTSTTTTTTCTTTTACCT
WI-12173	57 C	0.88 T	0.13	0.22	TACAAAAAATCCTGCYCTTATAGAGCATACA
WI-12179	96 G	0.50 A	0.50	0.50	GTACGGTGGAGGTCARGCATCTACAGGGTCA
WI-12201	61 C	0.69 T	0.31	0.43	CTGATCACCTGCATGYGCCAGGTATGTGGTC
WI-12210	76 A	0.88 G	0.13	0.22	AAACAACTATTGCATRGGAACATATGCAA
WI-12229	89 T	0.75 G	0.25	0.38	AAAAAGAGTAAAAATKACCAAAAAAATTAAG
WI-12234	66 A	0.44 G	0.56	0.49	ACACTTGTGGGGCTTCTTCAACATGGACTG

WI-12310	46 G	0.88 A	0.13	0.22	TAATTTTAAAAAGCTRTTTAGGACCCAAACA
WI-12319	109 T	0.88 C	0.13	0.22	GTTCTGCTCATAATTCYCAATATGTACCAGA
WI-12323	68 G	0.50 A	0.50	0.50	GTACCTATGAAATAARACAGGTAGGGAATAT
WI-12326	25 G	0.81 A	0.19	0.30	TCAAAAGCAATTCACRCTTCAGAAATACAAA
WI-12340	18 T	0.94 C	0.06	0.12	CAATATAATTCATTTCGAGTGATTAAACC
WI-12345	37 C	0.50 A	0.50	0.50	CAGGAAAAGAGGAAACCTGAACCCCTCTGC
WI-12361	63 C	0.00 T	1.00	0.00	CAGCATAATGATTAATYTGAACCTAAATTTACA
WI-12469	91 C	0.56 T	0.44	0.49	TATATTCTATTTCTAYTTTGACAGCACAGTTC
WI-12535	50 A	0.88 T	0.13	0.22	TTGAGGTGTAGATATWCTTCTCTCTCTCTCG
WI-12542	45 C	0.25 T	0.75	0.38	TGAACATTTAAATGTYATCCATGTGAGGGCT
WI-12542	70 G	0.50 T	0.50	0.50	AGGGCTCTAGATCATKGTAGGTGATTGATAC
WI-12542	71 G	0.63 T	0.38	0.47	GGGCTCTAGATCATKGTAGGTGATTGATACA
WI-12578	37 C	0.50 T	0.50	0.50	CTAAAGGAATGGGAAYGTGTGGTGGTGGCT
WI-12601	42 T	0.56 C	0.44	0.49	TATCTTGTCTTGATYGTCTACGTAAGCATG
WI-12634	52 T	0.31 C	0.69	0.43	TGTCTAGCAGTATTAYGCTATTAGCTATGTT
WI-12648	41 A	0.38 G	0.63	0.47	TGGCATTAAAGGATGCRGTAGGATGTCCACTT
WI-12684	64 G	0.19 T	0.81	0.30	TGTAAACAGCTGTGCKCCATTAGGCTTTGT
WI-12837	87 A	0.13 G	0.88	0.22	TCAAGGTAAAGTCCARTACAAAAAACAGCA
WI-12988	36 C	0.56 A	0.44	0.49	GTGCTCTCAGTACAAMAAACAGCATCAGTAG
WI-13020	108 G	0.81 A	0.19	0.30	AACCTGAGACTTTARATCTGCAAAGGGGTT
WI-13112	71 C	0.13 T	0.88	0.22	GACTTAAGCTTTTTTYCTTTTCCATATAAT
WI-13119	51 C	0.94 G	0.06	0.12	GACACAATCAAGACTSACAGTAGCCTCAACC
WI-13119	114 G	0.88 C	0.13	0.22	GGACTACAGGCATGTSACACCACACCTGGTT
WI-13264	25 G	0.31 A	0.69	0.43	AAGGCTCTGCCCCATRTATTTCCCGTCCTCC
WI-13364	35 A	0.38 G	0.63	0.47	TTTTTAGTAGAAGCRGGAACAGTTGTCAAT
WI-13367	84 C	0.44 G	0.56	0.49	GAAGACTCACCAAGASAGGGTGGGGTGGGGA
WI-13373	52 G	0.94 A	0.06	0.12	GAATAAACATCTCACRAACTGTGCTCCTAG
WI-13416	71 C	0.50 A	0.50	0.50	TGACAAGAACACATAMAAATATTGAAATAT
WI-13424	66 G	0.88 A	0.13	0.22	TTACCCCTATTTCTCRTAGACCCCTGGGGAGA
WI-13446	22 G	0.50 C	0.50	0.50	TTCTTCACTCATCASCCTTCTGATTTTGTAT

WI-13453	88 T	0.63 A	0.38	0.47	AAATCTGTCTCTCTCWTGCTAGAAAAGAGATG
WI-13470	100 C	0.81 A	0.19	0.30	ATATTGGAATTTCTAMAGAGACCCATGGTCT
WI-13473	31 C	0.94 T	0.06	0.12	ATGGGCTGAGACTGYTGCTGGTAGATGCA
WI-13477	32 A	0.44 G	0.56	0.49	TTGTTGGATAAAAAGGRCATTGTTTTTCATTA
WI-13477	61 A	0.88 G	0.13	0.22	TAGCTTGCTCTCAAARGACAGAGAAAATAAGA
WI-13507	41 T	0.94 C	0.06	0.12	AGCTTGACCTTAGGTYAATATTTCATTTGGG
WI-13522	33 C	0.31 T	0.69	0.43	CCCCACTAATACAAACYGAGAACCCACTGACTT
WI-13528	80 A	0.44 G	0.56	0.49	AAAAAGAAAGACATTTTTCAGAGAAAAACTGT
WI-13529	42 T	0.75 C	0.25	0.38	ATTGAACAGTTACCAYAAGCAAGAGAGAGTGAG
WI-13536	29 T	0.94 C	0.06	0.12	AAAAACTCAGCGAAGYGAAAAGGTGGATAGC
WI-13551	74 G	0.75 A	0.25	0.38	TATATTCAGACAATCRAATATTACTTAGCAC
WI-13578	48 T	0.63 A	0.38	0.47	AGCAGAAAAGAAAACCCWAGACAAAAAGATGTT
WI-13582	43 C	0.88 A	0.13	0.22	TCTAGAGACTGGGGAMTGGAATCTAACTGCG
WI-13594	66 G	0.75 A	0.25	0.38	CAGATCACAAAAGCRTCAGCACAAAAAAGTAC
WI-13600	26 G	0.88 T	0.13	0.22	GAGCCAAAGCATCCATKCCATCATCTAGTAAC
WI-13602	89 G	0.75 T	0.25	0.38	TCTGGAGACAACACAKAAATCTATTAATATT
WI-13650	76 A	0.56 T	0.44	0.49	TTTCACTTTAAAACWTAATAAAAACTACTCTT
WI-13654	49 A	0.63 G	0.38	0.47	TGAAACACATCCGTARGTATGACATCATTTTC
WI-13683	47 C	0.94 G	0.06	0.12	ACCTATCTGCCCATGSTTTTACAGCCTTTTAA
WI-13712	40 A	0.69 C	0.31	0.43	ATTTTTATTCTATTGMATTATAGAAAAAGTG
WI-13725	56 A	0.88 C	0.13	0.22	GCACATATGGGTGCCMGCCCCGAGACAGCAGG
WI-13744	115 C	0.38 T	0.63	0.47	CTGAACAAAACCTGAAYGCTGTGCTTATCTTT
WI-13752	106 T	0.81 C	0.19	0.30	AAGTGTGGATATACYTGGCTTGACCCGGAC
WI-13752	117 C	0.31 T	0.69	0.43	ATACTGGCTTGACACACACACCTTTTACGG
WI-13763	59 T	0.88 C	0.13	0.22	GGACACTGCAGTGATYAGGGGCAGGTGTGGG
WI-13785	27 T	0.31 C	0.69	0.43	ACTATAAAGTGCTTYAAAAATGCAGCAGCAG
WI-13785	40 C	0.38 G	0.63	0.47	TTTAAATGCAGCAGSAGGAGATGTGAAGAC
WI-13785	56 A	0.56 C	0.44	0.49	AGGAGATGTGAAGACMCAAAATGAACAAGTGC
WI-13785	72 G	0.56 A	0.44	0.49	CAAATGAACAAGTGCRTAGTGACACATAGCT
WI-13789	62 G	0.63 A	0.38	0.47	GGATGGCTGAGGGAGRGAACAGAGGAAGCGC

WI-13793	88 C		0.31 G	0.69	0.43	CAGCCTAGATATAGGSAGTAACAAATCCTCC
WI-13794	52 A		0.44 G	0.56	0.49	ACCTTTTCTTCTCTRTACAAAGTTAAGAGC
WI-13805	112 G		0.44 A	0.56	0.49	AAGGCACACGGGGAARGGTCAAGGCAGGCT
WI-13805	113 G		0.44 A	0.56	0.49	AGGCACACGGGGAARGGTCAAGGCAGGCTG
WI-13806	62 G		0.94 A	0.06	0.12	AACTAGGCCTCAGGTRCCCATTAAGCATGCT
WI-13810	106 T		0.81 C	0.19	0.30	ATACATCCAAAACCTTYAGTTAGCAGCAAGCA
WI-13831	56 G		0.94 C	0.06	0.12	AGTGACCTGGAAAASGAGATTCACATACTT
WI-13831	113 T		0.25 C	0.75	0.38	CTTCTCTTCTGTAGAYGTCTCCATGTTACAG
WI-13850	51 A		0.88 G	0.13	0.22	TTTTAACACAGCCATRTTACAAAACATTGTCA
WI-13857	28 A		0.94 G	0.06	0.12	AATGCTTTTCTGAACRTACATTTTAGGTATC
WI-13859	84 G		0.94 A	0.06	0.12	TGAAAAGGAAACTATRACAAAACAAAGTATATA
WI-13892	50 G		0.81 A	0.19	0.30	TTTTAAATAGAACARCTTTGATTTTAGTA
WI-13909	80 G		0.88 A	0.13	0.22	ACTCTCTTCAAACCTCRAATATCTTTTTCAGA
WI-13909	93 A		0.88 T	0.13	0.22	TCGAATATCTTTTTCWGAGATGCTAGCTAG
WI-13910	63 C		0.38 T	0.63	0.47	ACGTCCTTTGTGCTAYGTGATAGTGCTT
WI-13936	123 C		0.81 T	0.19	0.30	ATTCAATAGCCTATCYAACTCCATGTGGGAG
WI-13951	39 C		0.63 T	0.38	0.47	AAGTAATGAACAAAAYAGACCCAGATCAGA
WI-13951	88 G		0.63 C	0.38	0.47	GTTAAATCTGGAGCASAATTCAGCAGCAAAT
WI-13960	39 A		0.81 C	0.19	0.30	TTAAATACTGATAGAMGATGCAAAATTTGTCC
WI-13967	103 A		0.56 C	0.44	0.49	ACAAGGAAATAAAAAMCAGCTTTTAGGAGATG
WI-13983	52 G		0.75 A	0.25	0.38	CCACTCCTTAAACCTRCCACTGGGCTAAGAG
WI-14061	68 C		0.94 T	0.06	0.12	CCGTACATACCTTATYACCCATTTTCATCCAC
WI-14065	29 T		0.50 C	0.50	0.50	AGGTCAGAGGCAATTYGAGATCCAGATTCA
WI-14078	61 C		0.19 T	0.81	0.30	TTAGGAAGAGCAAGAYGCAGTAAGAGACATG
WI-14083	47 C		0.31 T	0.69	0.43	GCTTAAAAACAACACTYATTTGTTATTTTCA
WI-14085	31 A		0.13 G	0.88	0.22	TGTAAGAAGAAAAACRTAACTAGCAGTGAA
WI-14102	22 C		0.50 A	0.50	0.50	AAACAAAGCAGAAAAAMCCACCATTAAACAAG
WI-14124	92 A		0.94 G	0.06	0.12	CGTTAACACTAAGCCRTATTATTCAAAAATGT
WI-14125	88 C		0.63 T	0.38	0.47	ATTTTGGACGACTAYGTGGCCATGCCATTC
WI-14136	120 G		0.75 A	0.25	0.38	ACCATGTCTTCACATRGCCCAAGAGACAGA



WI-14138	23 C	0.88 T	0.13	0.22	GGCACCAGAAAAGCTYATGTTCTATGTTATG
WI-14149	83 C	0.94 T	0.06	0.12	TTAGCGTTAAAGGAGYTGAGTTGAGTCAAAC
WI-14153	28 A	0.56 G	0.44	0.49	TGCAGGAAGGCCAGCRTCCTCCTGCCGTT
WI-14162	57 A	0.81 G	0.19	0.30	TGGCCTCGCTGCCTCRGCCCTTTCTCTTTGA
WI-14186	52 C	0.50 T	0.50	0.50	ATGGAAGACACATAYGGTACAAAAATTACAG
WI-14186	88 A	0.50 G	0.50	0.50	TTAGTTCATTACATGRTACAAATCATTAGAG

### Analysis of Polymorphisms

#### A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source.

Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. See generally *PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila et al., *Nucleic Acids Res.* 19, 4967 (1991); Eckert et al., *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson et al., IRL Press, Oxford); and U.S. Patent 4,683,202 (each of which is incorporated by reference for all purposes).

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren et al., *Science* 241, 1077 (1988), transcription amplification (Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli et al., *Proc. Nat. Acad. Sci. USA*, 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

#### B. Detection of Polymorphisms in Target DNA

There are two distinct types of analysis depending whether a polymorphism in question has already been characterized. The first type of analysis

-33-

is sometimes referred to as de novo characterization. This analysis compares target sequences in different individuals to identify points of variation, i.e., polymorphic sites. By analyzing a groups of individuals representing the greatest ethnic diversity among humans and greatest breed and species variety in plants and animals, patterns characteristic of the most common alleles/haplotypes of the locus can be identified, and the frequencies of such populations in the population determined. Additional allelic frequencies can be determined for subpopulations characterized by criteria such as geography, race, or gender. The de novo identification of the polymorphisms of the invention is described in the Examples section. The second type of analysis is determining which form(s) of a characterized polymorphism are present in individuals under test. There are a variety of suitable procedures, which are discussed in turn.

#### 1. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki et al., *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15 mer at the 7 position; in a 16 mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

-34-

## 2. Tiling Arrays

The polymorphisms can also be identified by hybridization to nucleic acid arrays, some example of which are described by WO 95/11995 (incorporated by reference in its entirety for all purposes). One form of such arrays is described in the

5 Examples section in connection with de novo identification of polymorphisms. The same array or a different array can be used for analysis of characterized polymorphisms. WO 95/11995 also describes subarrays that are optimized for detection of a variant forms of a precharacterized polymorphism. Such a subarray contains probes designed to be complementary to a second reference sequence, which

10 is an allelic variant of the first reference sequence. The second group of probes is designed by the same principles as described in the Examples except that the probes exhibit complementarity to the second reference sequence. The inclusion of a second group (or further groups) can be particular useful for analyzing short subsequences of the primary reference sequence in which multiple mutations are expected to occur

15 within a short distance commensurate with the length of the probes (*i.e.*, two or more mutations within 9 to 21 bases).

## 3. Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer

20 exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers leading to a detectable product signifying the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the

25 polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer. See, *e.g.*, WO 93/22456.

-35-

#### 4. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind et al., *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

#### 5. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., *PCR Technology, Principles and Applications for DNA Amplification*, (W.H. Freeman and Co, New York, 1992), Chapter 7.

#### 6. Single-Strand Conformation Polymorphism Analysis

Alleles of target sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita et al., *Proc. Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence difference between alleles of target sequences.

### III. Methods of Use

After determining polymorphic form(s) present in an individual at one or more polymorphic sites, this information can be used in a number of methods.

-36-

A. Forensics

Determination of which polymorphic forms occupy a set of polymorphic sites in an individual identifies a set of polymorphic forms that distinguishes the individual. *See generally* National Research Council, *The Evaluation of Forensic DNA Evidence* (Eds. Pollard et al., National Academy Press, DC, 1996). The more sites that are analyzed the lower the probability that the set of polymorphic forms in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked. Thus, polymorphisms of the invention are often used in conjunction with polymorphisms in distal genes. Preferred polymorphisms for use in forensics are diallelic because the population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci.

The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (e.g., by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance.

$p(\text{ID})$  is the probability that two random individuals have the same polymorphic or allelic form at a given polymorphic site. In diallelic loci, four genotypes are possible: AA, AB, BA, and BB. If alleles A and B occur in a haploid genome of the organism with frequencies  $x$  and  $y$ , the probability of each genotype in a diploid organism are (see WO 95/12607):

Homozygote:  $p(\text{AA}) = x^2$

$$p(\text{ID}) = (x^2)^2 + (2xy)^2 + (y^2)^2.$$
$$p(\text{ID}) = x^4 + (2xy)^2 + (2yz)^2 + (2xz)^2 + z^4 + y^4$$

15           The cumulative probability of identity (cum p(ID)) for each of multiple  
unlinked loci is determined by multiplying the probabilities provided by each locus.

$$\text{cum } p(\text{ID}) = p(\text{ID1})p(\text{ID2})p(\text{ID3}).... p(\text{IDn})$$

$$\text{cum } p(\text{nonID}) = 1 - \text{cum } p(\text{ID}).$$

### B. Paternity Testing

The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing

-38-

investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child.

If the set of polymorphisms in the child attributable to the father does not match the putative father, it can be concluded, barring experimental error, that the putative father is not the real father. If the set of polymorphisms in the child attributable to the father does match the set of polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of coincidental match.

The probability of parentage exclusion (representing the probability that a random male will have a polymorphic form at a given polymorphic site that makes him incompatible as the father) is given by the equation (see WO 95/12607):

$$p(\text{exc}) = xy(1-xy)$$

where  $x$  and  $y$  are the population frequencies of alleles A and B of a diallelic polymorphic site.

(At a triallelic site  $p(\text{exc}) = xy(1-xy) + yz(1-yz) + xz(1-xz) + 3xyz(1-xyz)$ ), where  $x$ ,  $y$  and  $z$  are the respective population frequencies of alleles A, B and C).

The probability of non-exclusion is

$$p(\text{non-exc}) = 1 - p(\text{exc})$$

The cumulative probability of non-exclusion (representing the value obtained when  $n$  loci are used) is thus:

$$\text{cum } p(\text{non-exc}) = p(\text{non-exc1})p(\text{non-exc2})p(\text{non-exc3})\dots p(\text{non-excn})$$

The cumulative probability of exclusion for  $n$  loci (representing the probability that a random male will be excluded)

$$\text{cum } p(\text{exc}) = 1 - \text{cum } p(\text{non-exc}).$$

If several polymorphic loci are included in the analysis, the cumulative probability of exclusion of a random male is very high. This probability can be taken into account in assessing the liability of a putative father whose polymorphic marker set matches the child's polymorphic marker set attributable to his/her father.



-39-

C. Correlation of Polymorphisms with Phenotypic Traits

The polymorphisms of the invention may contribute to the phenotype of an organism in different ways. Some polymorphisms occur within a protein coding sequence and contribute to phenotype by affecting protein structure. The effect may be neutral, beneficial or detrimental, or both beneficial and detrimental, depending on the circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on replication, transcription, and translation. A single polymorphism may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by polymorphisms in different genes. Further, some polymorphisms predispose an individual to a distinct mutation that is causally related to a certain phenotype.

Phenotypic traits include diseases that have known but hitherto unmapped genetic components (e.g., agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial hypercholesterolemia, polycystic kidney disease, hereditary spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary hemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos syndrome, osteogenesis imperfecta, and acute intermittent porphyria). Phenotypic traits also include symptoms of, or susceptibility to, multifactorial diseases of which a component is or may be genetic, such as autoimmune diseases, inflammation, cancer, diseases of the nervous system, and infection by pathogenic microorganisms. Some examples of autoimmune diseases include rheumatoid arthritis, multiple sclerosis, diabetes (insulin-dependent and non-independent), systemic lupus erythematosus and Graves disease. Some examples of cancers include cancers of the bladder, brain, breast, colon, esophagus, kidney, leukemia, liver, lung, oral cavity, ovary, pancreas, prostate, skin, stomach and uterus. Phenotypic traits also include characteristics such as longevity, appearance (e.g., baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments.

Correlation is performed for a population of individuals who have been

-40-

tested for the presence or absence of a phenotypic trait of interest and for polymorphic markers sets. To perform such analysis, the presence or absence of a set of polymorphisms (i.e. a polymorphic set) is determined for a set of the individuals, some of whom exhibit a particular trait, and some of which exhibit lack of the trait. The alleles of each polymorphism of the set are then reviewed to determine whether the presence or absence of a particular allele is associated with the trait of interest. Correlation can be performed by standard statistical methods such as a  $\chi$ -squared test and statistically significant correlations between polymorphic form(s) and phenotypic characteristics are noted. For example, it might be found that the presence of allele A1 at polymorphism A correlates with heart disease. As a further example, it might be found that the combined presence of allele A1 at polymorphism A and allele B1 at polymorphism B correlates with increased milk production of a farm animal.

Such correlations can be exploited in several ways. In the case of a strong correlation between a set of one or more polymorphic forms and a disease for which treatment is available, detection of the polymorphic form set in a human or animal patient may justify immediate administration of treatment, or at least the institution of regular monitoring of the patient. Detection of a polymorphic form correlated with serious disease in a couple contemplating a family may also be valuable to the couple in their reproductive decisions. For example, the female partner might elect to undergo in vitro fertilization to avoid the possibility of transmitting such a polymorphism from her husband to her offspring. In the case of a weaker, but still statistically significant correlation between a polymorphic set and human disease, immediate therapeutic intervention or monitoring may not be justified. Nevertheless, the patient can be motivated to begin simple life-style changes (e.g., diet, exercise) that can be accomplished at little cost to the patient but confer potential benefits in reducing the risk of conditions to which the patient may have increased susceptibility by virtue of variant alleles. Identification of a polymorphic set in a patient correlated with enhanced receptiveness to one of several treatment regimes for a disease indicates that this treatment regime should be followed.

For animals and plants, correlations between characteristics and phenotype are useful for breeding for desired characteristics. For example, Beitz et

-41-

al., US 5,292,639 discuss use of bovine mitochondrial polymorphisms in a breeding program to improve milk production in cows. To evaluate the effect of mtDNA D-loop sequence polymorphism on milk production, each cow was assigned a value of 1 if variant or 0 if wildtype with respect to a prototypical mitochondrial DNA sequence at each of 17 locations considered. Each production trait was analyzed individually with the following animal model:

$$Y_{ijkpn} = \mu + YS_i + P_j + X_k + \beta_1 + \dots \beta_{17} + PE_n + a_n + e_p$$

where  $Y_{ijkpn}$  is the milk, fat, fat percentage, SNF, SNF percentage, energy concentration, or lactation energy record;  $\mu$  is an overall mean;  $YS_i$  is the effect common to all cows calving in year-season;  $X_k$  is the effect common to cows in either the high or average selection line;  $\beta_1$  to  $\beta_{17}$  are the binomial regressions of production record on mtDNA D-loop sequence polymorphisms;  $PE_n$  is permanent environmental effect common to all records of cow  $n$ ;  $a_n$  is effect of animal  $n$  and is composed of the additive genetic contribution of sire and dam breeding values and a Mendelian sampling effect; and  $e_p$  is a random residual. It was found that eleven of seventeen polymorphisms tested influenced at least one production trait. Bovines having the best polymorphic forms for milk production at these eleven loci are used as parents for breeding the next generation of the herd.

#### D. Genetic Mapping of Phenotypic Traits

The previous section concerns identifying correlations between phenotypic traits and polymorphisms that directly or indirectly contribute to those traits. The present section describes identification of a physical linkage between a genetic locus associated with a trait of interest and polymorphic markers that are not associated with the trait, but are in physical proximity with the genetic locus responsible for the trait and co-segregate with it. Such analysis is useful for mapping a genetic locus associated with a phenotypic trait to a chromosomal position, and thereby cloning gene(s) responsible for the trait. See Lander et al., *Proc. Natl. Acad. Sci. (USA)* 83, 7353-7357 (1986); Lander et al., *Proc. Natl. Acad. Sci. (USA)* 84, 2363-2367 (1987); Donis-Keller et al., *Cell* 51, 319-337 (1987); Lander et al.,

-42-

*Genetics* 121, 185-199 (1989)). Genes localized by linkage can be cloned by a process known as directional cloning. See Wainwright, *Med. J. Australia* 159, 170-174 (1993); Collins, *Nature Genetics* 1, 3-6 (1992) (each of which is incorporated by reference in its entirety for all purposes).

5                   Linkage studies are typically performed on members of a family. Available members of the family are characterized for the presence or absence of a phenotypic trait and for a set of polymorphic markers. The distribution of polymorphic markers in an informative meiosis is then analyzed to determine which polymorphic markers co-segregate with a phenotypic trait. See, e.g., Kerem et al., *Science* 245,  
10   1073-1080 (1989); Monaco et al., *Nature* 316, 842 (1985); Yamoka et al., *Neurology* 40, 222-226 (1990); Rossiter et al., *FASEB Journal* 5, 21-27 (1991).  
Linkage is analyzed by calculation of LOD (log of the odds) values. A lod value is the relative likelihood of obtaining observed segregation data for a marker and a genetic locus when the two are located at a recombination fraction  $\theta$ , versus the situation in  
15   which the two are not linked, and thus segregating independently (Thompson & Thompson, *Genetics in Medicine* (5th ed, W.B. Saunders Company, Philadelphia, 1991); Strachan, "Mapping the human genome" in *The Human Genome* (BIOS Scientific Publishers Ltd, Oxford), Chapter 4). A series of likelihood ratios are calculated at various recombination fractions ( $\theta$ ), ranging from  $\theta = 0.0$  (coincident  
20   loci) to  $\theta = 0.50$  (unlinked). Thus, the likelihood at a given value of  $\theta$  is: probability of data if loci linked at  $\theta$  to probability of data if loci unlinked. The computed likelihoods are usually expressed as the  $\log_{10}$  of this ratio (i.e., a lod score). For example, a lod score of 3 indicates 1000:1 odds against an apparent observed linkage being a coincidence. The use of logarithms allows data collected from different  
25   families to be combined by simple addition. Computer programs are available for the calculation of lod scores for differing values of  $\theta$  (e.g., LIPED, MLINK (Lathrop, *Proc. Nat. Acad. Sci. (USA)* 81, 3443-3446 (1984)). For any particular lod score, a recombination fraction may be determined from mathematical tables. See Smith et al., *Mathematical tables for research workers in human genetics* (Churchill, London,  
30   1961); Smith, *Ann. Hum. Genet.* 32, 127-150 (1968). The value of  $\theta$  at which the lod

-43-

score is the highest is considered to be the best estimate of the recombination fraction.

Positive lod score values suggest that the two loci are linked, whereas negative values suggest that linkage is less likely (at that value of  $\theta$ ) than the possibility that the two loci are unlinked. By convention, a combined lod score of +3 or greater (equivalent to greater than 1000:1 odds in favor of linkage) is considered definitive evidence that two loci are linked. Similarly, by convention, a negative lod score of -2 or less is taken as definitive evidence against linkage of the two loci being compared. Negative linkage data are useful in excluding a chromosome or a segment thereof from consideration. The search focuses on the remaining non-excluded chromosomal locations.

#### IV. Modified Polypeptides and Gene Sequences

The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise one of the sequences described in Table 1, column 8, in which the polymorphic position is occupied by one of the alternative bases for that position. Some nucleic acid encode full-length variant forms of proteins. Similarly, variant proteins have the prototypical amino acid sequences of encoded by nucleic acid sequence shown in Table 1, column 8, (read so as to be in-frame with the full-length coding sequence of which it is a component) except at an amino acid encoded by a codon including one of the polymorphic positions shown in the Table. That position is occupied by the amino acid coded by the corresponding codon in any of the alternative forms shown in the Table.

Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host

-44-

sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, *e.g.*, mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide. Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like.

The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, *i.e.*, 80, 95 or 99% free of cell component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Scopes, *Protein Purification, Principles and Practice*, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), *Guide to Protein Purification, Methods in Enzymology*, Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the protein can be isolated from a lysate of the host cells.

The invention further provides transgenic nonhuman animals capable of expressing an exogenous variant gene and/or having one or both alleles of an endogenous variant gene inactivated. Expression of an exogenous variant gene is usually achieved by operably linking the gene to a promoter and optionally an enhancer, and microinjecting the construct into a zygote. See Hogan et al., "Manipulating the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory. Inactivation of endogenous variant genes can be achieved by forming a transgene in which a cloned variant gene is inactivated by insertion of a positive selection marker. See Capecchi, *Science* 244, 1288-1292 (1989). The transgene is then

-45-

introduced into an embryonic stem cell, where it undergoes homologous recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

In addition to substantially full-length polypeptides expressed by variant genes, the present invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

Polyclonal and/or monoclonal antibodies that specifically bind to variant gene products but not to corresponding prototypical gene products are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles and Practice* (2d ed.) Academic Press, New York (1986). Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

## V. Kits

The invention further provides kits comprising at least one allele-specific oligonucleotide as described above. Often, the kits contain one or more pairs of allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a substrate. For example, the same substrate can comprise allele-specific oligonucleotide probes for detecting at least 10, 100 or all of the polymorphisms shown in Table 1. Optional additional components of the kit include, for example, restriction enzymes, reverse-

-46-

transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

### EXAMPLES

The polymorphisms shown in Table 1 were identified by resequencing of target sequences from eight unrelated individuals of diverse ethnic and geographic backgrounds by hybridization to probes immobilized to microfabricated arrays. The strategy and principles for design and use of such arrays are generally described in WO 95/11995. The strategy provides arrays of probes for analysis of target sequences showing a high degree of sequence identity to the reference sequences of the fragments shown in Table 1, column 1. The reference sequences were sequence-tagged sites (STSs) developed in the course of the Human Genome Project (*see, e.g., Science* 270, 1945-1954 (1995); *Nature* 380, 152-154 (1996)). Most STS's ranged from 100 bp to 300 bp in size.

A typical probe array used in this analysis has two groups of four sets of probes that respectively tile both strands of a reference sequence. A first probe set comprises a plurality of probes exhibiting perfect complementarity with one of the reference sequences. Each probe in the first probe set has an interrogation position that corresponds to a nucleotide in the reference sequence. That is, the interrogation position is aligned with the corresponding nucleotide in the reference sequence, when the probe and reference sequence are aligned to maximize complementarity between the two. For each probe in the first set, there are three corresponding probes from three additional probe sets. Thus, there are four probes corresponding to each nucleotide in the reference sequence. The probes from the three additional probe sets are identical to the corresponding probe from the first probe set except at the interrogation position, which occurs in the same position in each of the four corresponding probes from the four probe sets, and is occupied by a different nucleotide in the four probe sets. In the present analysis, probes were 25 nucleotides



-47-

long. Arrays tiled for multiple different references sequences were included on the same substrate.

Multiple target sequences from an individual were amplified from human genomic DNA using primers for the fragments indicated in the listed Web sites. The amplified target sequences were fluorescently labelled during or after PCR. The labelled target sequences were hybridized with a substrate bearing immobilized arrays of probes. The amount of label bound to probes was measured. Analysis of the pattern of label revealed the nature and position of differences between the target and reference sequence. For example, comparison of the intensities of four corresponding probes reveals the identity of a corresponding nucleotide in the target sequences aligned with the interrogation position of the probes. The corresponding nucleotide is the complement of the nucleotide occupying the interrogation position of the probe showing the highest intensity (see WO 95/11995). The existence of a polymorphism is also manifested by differences in normalized hybridization intensities of probes flanking the polymorphism when the probes hybridized to corresponding targets from different individuals. For example, relative loss of hybridization intensity in a "footprint" of probes flanking a polymorphism signals a difference between the target and reference (i.e., a polymorphism) (see EP 717,113, incorporated by reference in its entirety for all purposes). Additionally, hybridization intensities for corresponding targets from different individuals can be classified into groups or clusters suggested by the data, not defined *a priori*, such that isolates in a give cluster tend to be similar and isolates in different clusters tend to be dissimilar. See WO 97/29212 filed February 7, 1997 (incorporated by reference in its entirety for all purposes). Hybridizations to samples from different individuals were performed separately. Table 1 summarizes the data obtained for target sequences in comparison with a reference sequence for the eight individuals tested.

From the foregoing, it is apparent that the invention includes a number of general uses that can be expressed concisely as follows. The invention provides for the use of any of the nucleic acid segments described above in the diagnosis or monitoring of diseases, such as cancer, inflammation, heart disease, diseases of the CNS, and susceptibility to infection by microorganisms. The invention further

-48-

provides for the use of any of the nucleic acid segments in the manufacture of a medicament for the treatment or prophylaxis of such diseases. The invention further provides for the use of any of the DNA segments as a pharmaceutical.

5 All publications and patent applications cited above are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent application were specifically and individually indicated to be so incorporated by reference. Although the present invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent that certain changes and modifications may be practiced within the  
10 scope of the appended claims.

-49-

WHAT IS CLAIMED IS:

- 1                    1.     A nucleic acid segment of between 10 and 100 bases from a  
2 fragment shown in Table 1 including a polymorphic site, or the complement of the  
3 segment.
- 1                    2.     The nucleic acid segment of claim 1 that is DNA.
- 1                    3.     The nucleic acid segment of claim 1 that is RNA.
- 1                    4     The segment of claim 1 that is less than 50 bases.
- 1                    5.     The segment of claim 1 that is less than 20 bases.
- 1                    6.     The segment of claim 1, wherein the fragment is 19201 and the  
2 polymorphic site is at position 179.
- 1                    7.     The segment of claim 1, wherein the polymorphic site is  
2 diallelic.
- 1                    8.     The segment of claim 1, wherein the polymorphic form  
2 occupying the polymorphic site is the reference base for the fragment listed in Table  
3 1, column 3.
- 1                    9.     The segment of claim 1, wherein the polymorphic form  
2 occupying the polymorphic site is an alternative form for the fragment listed in Table  
3 1, column 5.
- 1                    10.    An allele-specific oligonucleotide that hybridizes to a segment  
2 of a fragment shown in Table 1, column 8 or its complement.

-50-

- 1                    11.    The allele-specific oligonucleotide of claim 10 that is probe.
- 1                    12.    The allele-specific oligonucleotide of claim 10, wherein a central  
2 position of the probe aligns with the polymorphic site of the fragment.
- 1                    13     The allele-specific oligonucleotide of claim 10 that is a primer.
- 1                    14.    The allele-specific oligonucleotide of claim 13, wherein the 3'  
2 end of the primer aligns with the polymorphic site of the fragment.
- 1                    15.    An isolated nucleic acid comprising a sequence of Table 1,  
2 column 8 or the complement thereof, wherein the polymorphic site within the sequence  
3 or complement is occupied by a base other than the reference base show in Table 1,  
4 column 3.
- 1                    16.    A method of analyzing a nucleic acid, comprising:  
2 obtaining the nucleic acid from an individual; and  
3 determining a base occupying any one of the polymorphic sites shown in Table  
4 1.
- 1                    17.    The method of claim 16, wherein the determining comprises  
2 determining a set of bases occupying a set of the polymorphic sites shown in Table 1.
- 1                    18.    The method of claim 16, wherein the nucleic acid is obtained  
2 from a plurality of individuals, and a base occupying one of the polymorphic positions  
3 is determined in each of the individuals, and the method further comprising testing  
4 each individual for the presence of a disease phenotype, and correlating the presence  
5 of the disease phenotype with the base.